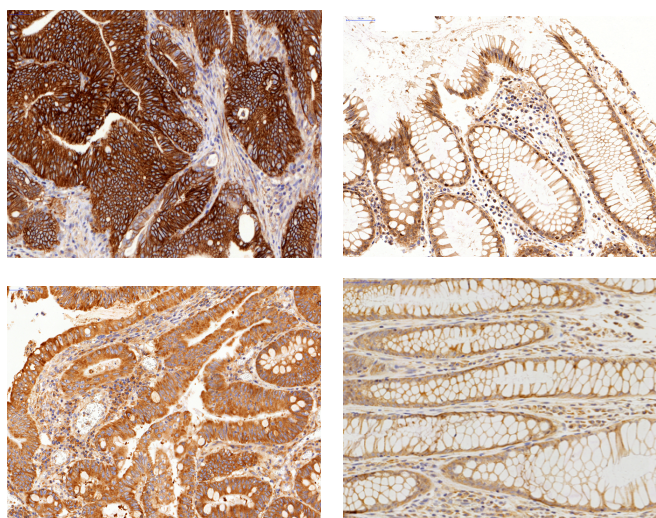


Departamento de Bioquímica
Facultad de Medicina
Universidad Autónoma de Madrid

Validación proteómica de dianas genómicas en cáncer colorrectal



PAULA LÓPEZ SERRA
TESIS DOCTORAL
Madrid, 2008

Departamento de Bioquímica
Facultad de Medicina

Universidad Autónoma de Madrid



Validación proteómica de dianas genómicas en cáncer colorrectal

Paula López Serra
Licenciatura en Biología

Director de la tesis: Dr. Jose Ignacio Casal Álvarez

Unidad de Tecnología de Proteínas
Centro Nacional de Investigaciones Oncológicas

El Dr. Jose Ignacio Casal Álvarez, jefe de la unidad de Tecnología de Proteínas del Centro Nacional de Investigaciones Oncológicas (CNIO) y profesor del Departamento de Bioquímica de la Facultad de Medicina de la Universidad Autónoma de Madrid,

CERTIFICA

Que el trabajo titulado *Validación proteómica de dianas genómicas en Cáncer colorrectal* que presenta Paula López Serra para la obtención del Grado de Doctor por la Universidad Autónoma de Madrid, ha sido realizado bajo mi dirección en la unidad de Tecnología de Proteínas del CNIO y reúne, a mi juicio, la calidad y contenidos suficientes para que sea presentado ante el tribunal correspondiente.

Para que así conste a los efectos oportunos, expide el presente certificado en Madrid a 12 de Junio de 2008.

Dr. Jose Ignacio Casal Álvarez

Esta tesis doctoral ha sido financiada por la Unidad de Tecnología de Proteínas del Centro Nacional de Investigaciones Oncológicas (CNIO) y por una Beca de Formación en Investigación Predoctoral (BEFI) otorgada por el Instituto de Salud Carlos III, Ministerio de Ciencia y Consumo de España.

A Lidia

Resumen en castellano

El cáncer colorrectal es la principal causa de muerte por cáncer en los países desarrollados. Dada la elevada mortalidad asociada a este tipo de cáncer es necesario un continuo descubrimiento de nuevas dianas terapéuticas que permitan disponer de nuevos y mejores tratamientos.

El adenocarcinoma surge en las células epiteliales del colon, primero se genera una masa de glándulas anormales que invaden y finalmente penetran los tejidos adyacentes. Una vez formados, los adenomas pueden continuar creciendo hasta que la membrana situada debajo del adenoma se rompe y el adenocarcinoma se vuelve invasivo.

Desde un punto de vista molecular, el CCR es uno de los tumores sólidos más estudiados. La mayor parte de estos estudios se centran en alteraciones genéticas, sin embargo, pocos han analizado la función e importancia de los genes individualmente y su correlación entre la expresión génica de los genes alterados en cáncer y la expresión de sus productos proteicos. Nuestro estudio parte del análisis de la expresión génica diferencial entre 22 pares de muestras normales y tumorales de pacientes de CCR utilizando *microarrays* de cDNA. Se seleccionaron un total de 29 genes a partir de los 50 con mayor grado de sobreexpresión para obtener sus productos proteicos y anticuerpos específicos frente a ellos, necesarios para caracterizar a nivel tisular la expresión proteica diferencial entre tejidos sanos y tumorales.

Se utilizó el sistema *Gateway* y *Escherichia coli* para la expresión de las proteínas seleccionadas. En nuestro estudio, se utilizaron tres *tags* diferentes con cada proteína para aumentar las posibilidades de obtenerlas como productos solubles, más adecuados para la producción de anticuerpos. Una vez obtenidas las proteínas seleccionamos anticuerpos recombinantes mediante la tecnología del despliegue en fagos, anticuerpos monoespecíficos mediante la inmunización de conejos con PrEST y anticuerpos mono y policlonales mediante los métodos clásicos del hibridoma e inmunización de conejos, respectivamente.

La validación de los genes seleccionados a nivel proteico mediante técnicas inmunológicas (*western blot* e inmunohistoquímica) confirmó que existe una buena correlación entre las alteraciones génicas y proteicas. Sin embargo, los estudios comparativos entre líneas celulares y muestras tumorales reflejaron que una gran parte de las dianas analizadas provienen principalmente de alteraciones en el estroma tumoral, no de las células tumorales propiamente dichas.

English abstract

Colorectal cancer (CRC) is the leading cause of cancer mediated deaths in developed countries. Due to its high mortality rate it is necessary to find new therapeutic targets and develop new treatment. The adenocarcinoma begins in the colonic epithelial cells. Abnormal glands appear and invade proximal locations. These adenomas can grow and achieve the underlying membrane becoming invasive.

From a molecular point of view, CRC is one of the best studied solid tumors. Most of the studies are focused in genetic alterations, however, only a few of them have pursued further insight into the function and/or importance of individual genes and the correlation between genomic alterations and their protein profile. We used cDNA microarrays to identify differentially expressed genes in malignant *versus* normal samples isolated from 22 individual patients with CRC. 29 genes were selected from the 50 with the highest expression for protein products and specific antibodies production in order to analyze the differential proteomic profile in normal and tumoral tissues.

Gateway system and *Escherichia coli* were selected for the protein expression. In our study, we used three different tags for each protein in order to obtain them as soluble products. Recombinant antibodies against the selected proteins were obtained by phage display, monospecific antibodies were obtained by rabbit immunization with PrEST and mono and polyclonal antibodies were selected by hybridoma technology and rabbit immunization, respectively.

Proteomic validation of the selected genes by immunological methods (western blot and immunohistochemistry) confirmed a good correlation between the transcriptomic analysis and the protein expression data. However, comparative studies between human cancer cell lines and tumoral samples showed that most of the analyzed genes are altered in the tumoral stroma but not in the tumoral cells.

Clave de abreviaturas

CCR	Cáncer colorrectal
CDR	Región determinante de complementariedad
CNIO	Centro Nacional de Investigaciones Oncológicas
cDNA	DNA complementario
2D-DIGE	<i>Two-dimensional differential in gel electrophoresis</i>
DNA	Ácido desoxirribonucleico
DO	Densidad óptica
DTT	Ditiotreitol
ELISA	<i>Enzyme-linked immunosorbent assay</i>
GFP	Proteína verde fluorescente
GST	Glutation S Transferasa
6xHis	Cola de 6 histidinas
IgG	Inmunoglobulina G
IgM	Inmunoglobulina M
IMAGE	<i>Integrated molecular análisis of genome expression consortium</i>
IMAC	<i>Immobilized metal ion affinity chromatography</i>
IPTG	Isopropyl-1-thio- β -Dgalactopiranosido
K _{on}	Constante de asociación
K _{off}	Constante de disociación
K _D	Constante de afinidad
LB	Medio Luria-Bertani
LC-ESI-MS/MS	<i>Liquid-chromatography-ion-spray tandem mass spectrometry</i>
mAb	Anticuerpo monoclonal
MALDI-TOF	<i>Matrix-assisted laser-desorption-ionization time-of-flight</i>
MBP	Proteína de unión a la maltosa
MGC	<i>Mammalian Gene Collection</i>
mRNA	RNA mensajero
msAb	Anticuerpo monoespecífico
ORF	Fase de lectura abierta
pAb	Anticuerpo policlonal
PBS	Tampón fosfato salino

PCR	Reacción en cadena de la polimerasa
PEG	Polietilenglicol
PrEST	<i>Protein epitope signature tag</i>
PSA	Persulfato amónico (en geles de poliacrilamida)
PSA	Antígeno específico de próstata (marcador tumoral)
RNA	Ácido ribonucleico
RU	Unidades de resonancia
rpm	revoluciones por minuto
SAM	<i>Significance Anaisys of Microarrays</i>
scFv	<i>Single chain variable fragment</i>
SDS	Dodecil sulfato sódico
SDS-PAGE	Eletoforesis en gel de acrilamida con SDS
siRNA	<i>small interfering RNA</i>
SPR	Resonancia de plasmón en superficie
TMB	3,3',5, 5'-tetramethylbenzidina
Trx	Tiorredoxina
V _H	Región variable de la cadena pesada
V _L	Región variable de la cadena ligera

Índice

Introducción	1
1. El cáncer	3
2. Oncogénesis y carcinogénesis	3
3. Progresión tumoral y metástasis	6
4. El cáncer colorrectal	7
4.1 Clasificación del cáncer colorrectal	9
4.2 Alteraciones genéticas en el cáncer colorrectal	10
5. Marcadores tumorales	12
5.1 Marcadores en cáncer colorrectal	14
6. Sistemas de expresión de proteínas	14
6.1 Expresión de proteínas en <i>E. coli</i>	15
6.2 Sistemas de expresión de alto rendimiento.	
Tecnología Gateway	16
7. Anticuerpos en la investigación y la terapia del cáncer	18
8. Anticuerpos monoclonales	19
9. Tecnología del despliegue en fagos	20
10. Librerías de anticuerpos desplegados en fagos	22
11. Estudios de afinidad. Uso del Biacore	24
12. <i>Protein Epitope Signature Tag</i>	24
13. Validación de dianas genómicas a nivel proteómico	25
Objetivos	27
Materiales y Métodos	31
1. Cultivo de líneas celulares	33
2. Uso de <i>microarrays</i> de cDNA en muestras de cáncer colorectal	33
3. PCR cuantitativa en tumores y líneas celulares	34
3.1 Extracción de RNA total y síntesis de cDNA a partir de tejido y líneas celulares	34
3.2 PCR Cuantitativa	35
4. Expresión de proteínas	35
4.1 Predicción de regiones transmembrana y péptidos señal	35
4.2 Obtención de los clones de cDNA	36
4.3 Diseño de oligonucleótidos	36
4.4 Clonaje en el sistema pENRT/D-TOPO	39
5. Clonaje en los vectores destino	39
6. Expresión de proteínas en <i>E. coli</i>	40
7. Electroforesis en geles de poliacrilamida y <i>western blot</i>	41
8. Purificación de las proteínas recombinantes	42
8.1 Purificación de proteínas fusionadas a MBP	42
8.2 Purificación de proteínas fusionadas a GST	42
8.3 Purificación de proteínas fusionadas a6xHis	43
8.4 Replegamiento de las proteínas insolubles	43

9. Producción de anticuerpos recombinantes mediante despliegue en fagos.....	44
9.1 Obtención de anticuerpos a partir de las librerías Tomlinson.....	44
9.1.1 Preparación de las librerías.....	44
9.1.2 Titulación del fago.....	44
9.1.3 Selección de scFv.....	45
9.1.4 ELISA de fagos policlonales.....	46
9.1.5 ELISA de fagos monoclonales.....	46
9.1.6 ELISA de scFv.....	46
9.1.7 Análisis de las secuencias.....	47
9.1.8 <i>Western blot</i> con las scFv seleccionadas.....	48
9.1.9 Eliminación del codon <i>stop</i> TAG en las scFv.....	48
9.1.10 Expresión de las scFv sin pIII en TG1.....	50
9.2 Obtención de anticuerpos a partir de las librerías Mehta.....	50
9.2.1 Preparación de las librerías Mehta.....	50
9.2.2 Selección de scFv.....	50
9.2.3 Expresión y purificación de las scFv Mehta en pET28b.....	51
9.2.4 Expresión y purificación de las scFv en pSANG10-3F.....	52
9.2.5 ELISA de las scFv fusionadas a 3xFLAG.....	53
9.2.6 Expresión y purificación de las scFv fusionadas al 3xFLAG en pET32b.....	54
10. Análisis de las constantes cinéticas de las scFv seleccionadas.....	54
11. Obtención de anticuerpos monoespecíficos frente a fragmentos proteicos (PrEST).....	56
11.1 Diseño de los fragmentos proteicos PrEST.....	56
11.2 Expresión y purificación de fragmentos PrEST.....	56
11.3 Obtención de sueros policlonales.....	58
11.4 ELISA de sueros policlonales.....	58
11.5 Purificación de anticuerpos monoespecíficos.....	59
12. Producción de anticuerpos monoclonales.....	59
13. <i>Western blot</i> con muestras de tejido.....	60
14. Preparación y análisis de <i>microarrays</i> de tejido.....	61
14.1 Tinción con scFv.....	61
14.2 Tinción con anticuerpos monoespecíficos.....	62
14.3 Tinción con anticuerpos mono y policlonales.....	62
15. Transfección de siRNA.....	62
16. Determinación de la apoptosis celular.....	63
17. Ensayo de proliferación celular.....	64
17.1 Construcción de los vectores pEGFP/C1-ANXA3/BMP4.....	64
17.2 Ensayo de proliferación celular.....	64

Resultados.....	67
1. Identificación de dianas moleculares de cáncer colorrectal.....	69
1.1 Identificación de genes sobreexpresados en cáncer colorrectal mediante el análisis de microarrays de DNA.....	69
1.2 Selección de proteínas diana a partir de genes alterados en cáncer colorrectal.....	69
1.3 Validación por PCR cuantitativa de las dianas seleccionadas.....	70
2. Producción de proteínas recombinantes a partir de los genes seleccionados...	72
2.1 Expresión de proteínas a partir de clones de cDNA de secuencia parcial.....	73
2.2 Expresión de proteínas a partir de clones de cDNA de secuencia completa.....	75
2.3 Purificación de proteínas recombinantes.....	79
3. Obtención de anticuerpos para la validación de las dianas Seleccionadas.....	80
3.1 Obtención de anticuerpos recombinantes humanos mediante despliegue en fagos (<i>Phage display</i>).....	81
3.1.1 Selección de scFv frente a las proteínas CXCL3, MMP7 y SPARC a partir de las librerías Tomlinson.....	82
3.1.1.1 Reactividad de las scFv seleccionadas.....	83
3.1.1.2 Secuenciación y caracterización de las scFv seleccionadas...	83
3.1.1.3 Preparación de scFv solubles.....	85
3.1.2 Selección de fagos a partir de la librería Mehta.....	86
3.1.2.1 Especificidad de las scFv seleccionadas.....	88
3.1.2.2 Secuenciación de las scFv seleccionadas.....	88
3.1.2.3 Expresión de las scFv Metha conteniendo 6xHis.....	90
3.1.2.4 Aplicación de las scFv al análisis de la expresión de proteínas diana en líneas celulares de cáncer colorectal.....	92
3.1.2.5 Expresión y purificación de las scFv seleccionadas con 3xFLAG.....	93
3.1.2.6 Expresión y purificación de las scFv fusionadas a 3xFLAG y tiorredoxina.....	94
3.1.3 Estudio de la afinidad de las scFv seleccionadas mediante resonancia del plasmón de superficie.....	96
3.2 Anticuerpos monoespecíficos basados en PrEST	98
3.2.1 Expresión y purificación de los fragmentos proteicos (PrEST) en <i>E. coli</i>	99
3.2.2 Obtención de anticuerpos monoespecíficos frente a PrEST.....	101
3.3 Anticuerpos monoclonales y policlonales.....	103
4. Validación de las dianas seleccionadas en CCR.....	104
4.1 Análisis de los niveles de expresión proteica en cáncer colorrectal mediante <i>western blot</i>	104

4.2 Análisis de los niveles de expresión proteica en cáncer colorrectal mediante inmunohistoquímica.....	106
5. Validación funcional de las dianas genómicas seleccionadas en CCR.....	109
5.1 Estudio de la apoptosis tras la interferencia de los genes ANXA3, BMP4 y MMP11 en las líneas celulares de cáncer de colon Hct15 y Sw480.....	109
5.2 Transfección de la línea celular Hct116 con los genes ANXA3 y BMP4 fusionados a GFP.....	111
Discusión	113
Conclusiones	115
Bibliografía	119
Anexo I	141
Anexo II	153

Introducción

1. El cáncer

El cáncer es un conjunto de enfermedades caracterizado por la presencia de un grupo de células que se dividen sin control, invaden el tejido en el que se encuentran y se diseminan por el organismo a través del torrente sanguíneo y el sistema linfático.

Existen diferentes tipos de cáncer dependiendo del tejido en el que se desarrolla y las características tumorales. Las principales categorías en que se dividen los tumores son carcinoma (cáncer que se origina en la piel o mucosas que recubren los órganos internos; es el tipo más frecuente), sarcoma (cáncer que se origina en los tejidos conectivos o de soporte como el hueso, el cartílago, la grasa o los vasos sanguíneos), leucemia (cáncer que se origina en el tejido formador de células sanguíneas, como la médula ósea, y que da lugar a un gran número de células que invaden el torrente sanguíneo), linfoma y mieloma (cánceres que se originan en las células del sistema inmune).

2. Oncogénesis y carcinogénesis

El proceso de aparición de un tumor se denomina carcinogénesis. Para que la carcinogénesis tenga lugar tienen que darse una serie de alteraciones genéticas que pueden tardar años en aparecer. La aparición de un tumor pasa por diferentes fases que van otorgando a la célula la capacidad de escapar de los mecanismos de control que aseguran la homeostasis de un tejido. La primera fase se produce cuando diversos agentes cancerígenos actúan sobre una célula normal alterando su material genético generando una mutación. Una sola mutación no es suficiente para que se desarrolle un tumor pero puede ser el desencadenante del proceso carcinogénico. La condición indispensable para que se produzca un tumor es que esta primera mutación otorgue a la célula alterada la capacidad de dividirse de forma incontrolada, esta mutación se transmite a las células descendientes y como resultado se produce una acumulación de células que se multiplican a una velocidad ligeramente superior a la normal. Este proceso se denomina fase de iniciación tumoral. La alteración producida es irreversible pero todavía insuficiente para dar lugar a un tumor.

Si sobre las células iniciales actúan de nuevo y de forma repetida los agentes carcinógenos, la multiplicación celular comienza a ser más rápida y la probabilidad de que se produzcan nuevas mutaciones aumenta, este proceso se denomina fase de

promoción. Actualmente se conocen muchos factores que actúan en esta segunda fase como por ejemplo el tabaco, una alimentación inadecuada, el alcohol, etc.

Las células tumorales acumulan nuevas mutaciones, ya sea por la división del material genético sin los controles adecuados que aseguran la integridad de las copias o por la actuación sobre el material genético de agentes carcinógenos. Las anomalías en su crecimiento y comportamiento aumentan y adquieren la capacidad de invasión tanto a nivel local, infiltrándose en los tejidos que las rodean, como a distancia, originando las metástasis. Esta última fase se denomina fase de progresión.

Se postula que son necesarias al menos tres mutaciones génicas (Komarova *et al*, 2003) en oncogenes, genes supresores de tumores y/o genes de reparación del DNA para desarrollar un tumor sólido en un individuo adulto. En el caso de los tumores sanguíneos son necesarias solamente dos mutaciones debido a que la presencia de las células tumorales en el torrente sanguíneo acelera el proceso invasivo del tumor.

La teoría más aceptada sobre el origen del cáncer postula una composición monoclonal del mismo (Fearon *et al*, 1987), es decir, todas las células cancerosas provienen de una única célula original que en un determinado momento adquiere la capacidad tumoral. Si una sola célula es capaz de generar un cáncer, ésta debe de pasar sus propiedades a las células hijas por lo que la aberración debe de ser heredable. Esta aberración heredable puede proceder de mutaciones genéticas (es la propia secuencia de DNA la que posee la mutación) o epigenéticas (es el patrón de expresión génica el que se encuentra alterado). Estadísticamente, la mayoría de los tumores se originan por cambios genéticos, de hecho, muchos compuestos cancerígenos lo son por generar mutaciones en la secuencia del DNA. Sin embargo, las mutaciones presentes en las células tumorales no siempre se deben a la actuación de agentes externos, a veces se deben a la propia naturaleza de las células dado que la maquinaria de replicación del DNA y la maquinaria de reparación de errores en el DNA, presentes en todas las células para garantizar la integridad de su secuencia genética, introducen mutaciones debido a la tasa de error propia de estos dos sistemas.

Los genes cuyas mutaciones o deleciones pueden dar lugar al desarrollo de un cáncer son los denominados oncogenes y genes supresores de tumores. El primer oncogén fue descubierto en 1970 por el Dr. Steve Martin en su laboratorio de la Universidad de California (Berkeley). Se trataba del oncogén *src* introducido en su

forma activa en células de pollo por el retrovirus Sarcoma de Rous. En 1976 los doctores J. Michael Bishop y Harold E. Varmus, de la Universidad de California (San Francisco), demostraron que los oncogenes eran proto-oncogenes defectuosos presentes en la mayoría de los organismos; este descubrimiento les valió la obtención del premio Nobel de Medicina en 1989.

Los oncogenes son aquellos genes cuya ganancia de función origina el cáncer mientras que los genes supresores de tumores son aquellos en los que es la pérdida de función la que promueve el cáncer. La forma no tumoral de un oncogén se denomina proto-oncogén; mutaciones en estos proto-oncogenes o cambios en su patrón de expresión dan lugar a los oncogenes, que promueven la proliferación y crecimiento celulares. Uno de los primeros oncogenes descubiertos en el desarrollo del cáncer fue el oncogén *ras*; mutaciones en la familia del proto-oncogén *Ras* (*H-Ras*, *N-Ras* y *K-Ras*) son muy comunes en todos los tipos de cáncer, apareciendo en un porcentaje de entre un 20% y un 30% en todos los tumores humanos (Bos, 1989).

Los genes supresores de tumores dan lugar a proteínas anti-proliferativas y supresoras del crecimiento celular y mitosis. Generalmente estos genes se activan ante el estrés celular o daño en el DNA; su función es detener la progresión del ciclo celular para evitar que los daños de la célula se pasen a las células hijas. Uno de los genes supresores de tumores más estudiado es P53. La proteína p53 induce el paro del ciclo celular o la apoptosis en respuesta a un daño severo en el DNA regulando la transcripción de genes que detienen el crecimiento y la proliferación celulares.

Como ya hemos visto, las mutaciones en los genes supresores de tumores y en los proto-oncogenes pueden producirse durante la vida de un individuo por exposición a agentes cancerígenos o por la tasa de error de los mecanismos reguladores de la replicación y reparación del DNA, sin embargo, existe un pequeño porcentaje de tumores de carácter hereditario. En este caso, la mutación desencadenante del proceso tumoral está presente en todas las células del organismo y ha sido heredada a través de los gametos de uno de los progenitores.

Una sola mutación no es suficiente para originar un tumor, deben de producirse varias mutaciones acumulativas en el mismo linaje celular para que la célula adquiera las propiedades tumorogénicas. A lo largo de los sucesivos ciclos de división celular las células van acumulando mutaciones que les proporcionan ventajas con respecto al resto

de células que forman el tejido y las capacitan para sobrevivir en un ambiente tumoral donde escasean el oxígeno y los nutrientes mientras no tenga lugar la angiogénesis, que consiste en la formación de nuevos vasos sanguíneos que irrigan el tumor.

3. Progresión tumoral y metástasis

La difusión a lo largo del organismo a través de los vasos sanguíneos y linfáticos origina nuevos tumores del mismo tipo celular en otros tejidos del individuo en un proceso denominado metástasis. Para ello, las células tumorales tienen que abandonar el tumor primario, invadir el tejido local y los vasos sanguíneos o linfáticos atravesando el endotelio y establecer nuevas colonias en otros tejidos atravesando un nuevo endotelio. Una vez alcanzado el tejido distal deben de sobrevivir y proliferar allí. La capacidad de invadir el propio tejido y tejidos distantes es lo que caracteriza a un tumor maligno. En la figura 1 se representa esquemáticamente el proceso tumoral desde su generación en un tejido hasta la invasión de nuevos tejidos.

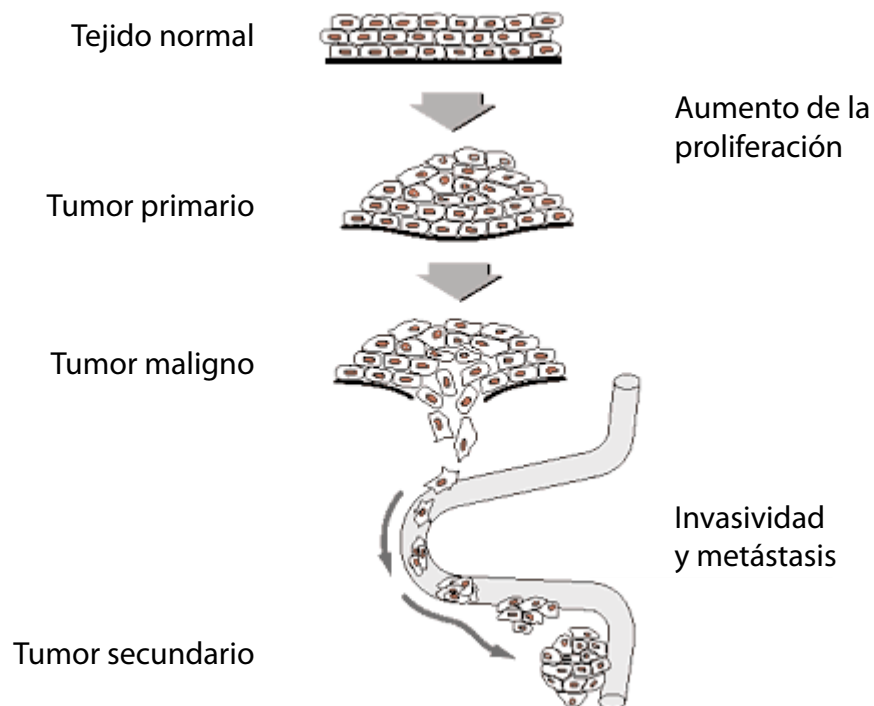


Figura 1. Progresión tumoral desde el tejido sano hasta la metástasis en tejidos distales tras invadir los vasos sanguíneos y/o linfáticos (Asociación Española Contra el Cáncer)

Para que un tumor pueda proliferar es necesario que disponga de oxígeno y nutrientes. Los tejidos sanos disponen de mecanismos para incrementar el riego sanguíneo cuando este es necesario en condiciones fisiológicas normales, los tumores

crecen rápidamente gracias al aumento de señales angiogénicas que promueven la formación de vasos sanguíneos que suplen al tumor de oxígeno y nutrientes. Estos vasos sanguíneos son además una vía de escape para las células metastásicas por lo que la angiogénesis es un proceso necesario tanto para el crecimiento de los tumores primarios como para la metástasis.

4. El cáncer colorrectal

Actualmente el cáncer es la principal causa de muertes por motivos naturales en España. Haciendo una clasificación por sexos, el cáncer de pulmón es el más frecuente en hombres y el de mama en mujeres, siendo el cáncer colorrectal el segundo tipo de cáncer más frecuente en ambos sexos. En términos absolutos el cáncer colorrectal es el cáncer más frecuente en España (Ministerio de Sanidad y Consumo de España).

El cáncer colorrectal es de uno de los tumores sólidos mejor estudiados desde un punto de vista molecular dado su fácil acceso, gran tamaño y alta incidencia (Fearon *et al*, 1996).

El cáncer colorrectal es un adenocarcinoma que se desarrolla en las células epiteliales de la mucosa glandular del colon y del recto. El colon está formado por aproximadamente 10^7 criptas compuestas por varios miles de células diferenciadas y entre 1 y 10 células madre. Las células madre están en el fondo de las criptas y se dividen lenta y asimétricamente mientras que las células diferenciadas se dividen rápidamente y se desplazan hasta el borde superior de la cripta, donde sufren apoptosis (Bach *et al*, 2000). Cada día se renuevan en el colon alrededor de 10^{10} células, lo que aumenta el riesgo de desarrollar cáncer debido a las mutaciones génicas que se producen durante la replicación del DNA y la segregación cromosómica (Rajagopalan *et al*, 2003).

En las etapas tempranas del desarrollo neoplásico una célula epitelial del colon adquiere una ventaja adaptativa sobre las células que la rodean y forma un tumor pequeño y premaligno. Las lesiones más tempranas aparecen en forma de glándulas hiperplásicas que forman criptas aberrantes en la mucosa. Las células hiperproliferativas se extienden a la superficie y se acumulan formando una masa celular llamada adenoma. Una vez formados, los adenomas continúan creciendo hasta convertirse en adenomas invasivos capaces de invadir el tejido que les rodea, expandiéndose tanto en

el interior del lumen intestinal como en el tejido muscular subyacente, invadiendo los vasos sanguíneos y linfáticos locales y sentando finalmente las bases para la metástasis final, preferentemente al hígado. En la Figura 2 se representa gráficamente la progresión tumoral del cáncer colorrectal.

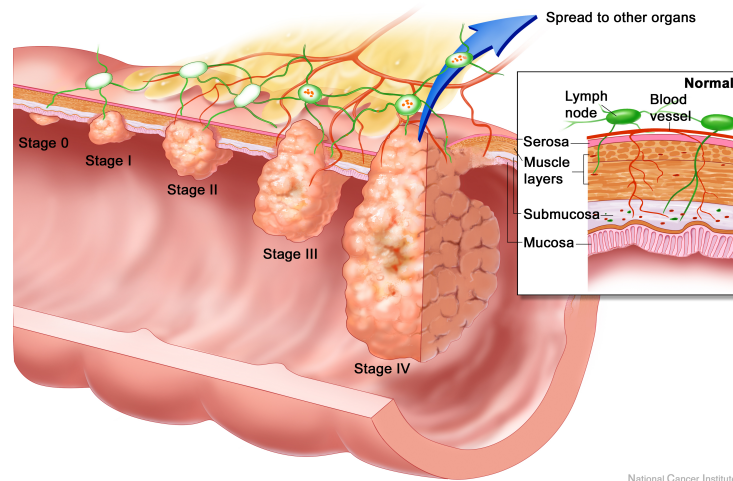


Figura 2. Progresión tumoral del cáncer colorrectal (U. S. National Cancer Institute).

Entre los factores ambientales más influyentes en el desarrollo del cáncer colorrectal se encuentran los hábitos alimenticios. Como factores de riesgo se han descrito el abuso del consumo de carnes rojas y grasas animales y la falta de consumo de fibras; por otro lado, se han señalado como factores protectores el consumo de vegetales, fruta, fibra y calcio. También se ha descrito que el consumo de tabaco aumenta la incidencia del cáncer colorrectal (Sharpe *et al*, 2002).

Se han definido dos tipos de cáncer colorrectal, esporádico y familiar. El cáncer colorrectal más frecuente es el de tipo esporádico (85% de los casos) en el que la neoplasia se debe a una mutación somática aberrante en el DNA de una célula en uno de sus ciclos de replicación. El cáncer familiar presenta componentes hereditarios, es decir, la mutación que genera la neoplasia está presente en la línea germinal de alguno de los progenitores y está presente en todas las células del organismo. El CCR hereditario agrupa el cáncer colorrectal hereditario no-polipósico o Síndrome de Lynch (5-10%), que se debe a la presencia de mutaciones o deleciones en uno o más genes reparadores de *mismatch* del DNA, como *hMLH1*, *hMSH2* o *hMSH6* y se caracteriza por presentar inestabilidad de microsatélites (Lynch *et al*, 1991), y la poliposis adenomatosa familiar (0.01%), en el que todas las células del organismo presentan una mutación en el gen

supresor de tumores *APC (Adenomatous polyposis coli) APC*, uno de los principales genes implicados en el desarrollo del cáncer colorrectal (Grodén *et al*, 1991). Por otro lado, se han descrito una serie de mutaciones en diferentes genes clave en la regulación de la señalización celular que aportan una mayor susceptibilidad a padecer cáncer colorrectal, entre estas mutaciones se incluyen el Síndrome Peutz-Jeghers, debido a la presencia de mutaciones germinales en el gen supresor de tumores *STK11* (o *LKB1*, serina-treonina quinasa que modula la proliferación y la polaridad celular) (Alessi *et al*, 2006), el Síndrome de la poliposis juvenil, debido a la presencia de mutaciones germinales en los genes *SMAD4*, *BMPR1A* y *ENG*, implicados en vía de señalización mediada por el *Transforming growth factor (TGF)-β* (Sweet *et al*, 2005), implicado en la proliferación, diferenciación y adhesión celulares, y el Síndrome PTEN, con alteraciones germinales en el gen supresor de tumores *PTEN (Phosphatase and tensin homologue)*, implicado en el crecimiento y proliferación celulares y la angiogénesis (Waite & Eng, 2002).

4.1 Clasificación del cáncer colorrectal

Se han hecho varias clasificaciones de los tumores colorrectales dependiendo de la gravedad de la neoplasia. Las clasificaciones más utilizadas son la de Dukes y el sistema TNM.

La clasificación de Dukes (Dukes, 1932) identifica diferentes estadios en la progresión tumoral: Dukes A cuando el tumor está definido en la pared intestinal, Dukes B si el tumor invade el tejido adyacente, Dukes C cuando el tumor ha invadido los ganglios linfáticos y Dukes D si hay metástasis al hígado. La supervivencia del paciente con CCR varía en gran medida dependiendo del estadio de progresión tumoral en el que se encuentra a la hora de diagnosticar la neoplasia siendo, por ejemplo, de un 84% en el caso del estadio Dukes B o de un 55% en el caso del estadio Dukes C (Mandel *et al*, 1999).

El sistema TNM hace referencia al grado de invasión de la pared intestinal (T), al grado de invasión de los ganglios linfáticos (N) y al grado de metástasis (M). Es el sistema de clasificación más utilizado en medicina aunque la clasificación de Dukes todavía se utiliza. El grado de invasión de la pared intestinal oscila entre 0 (no hay evidencias de neoplasia) y 4 (invasión de tejidos adyacentes, como el páncreas). En el caso de la invasión de ganglios linfáticos, el grado de invasividad oscila entre 0 (no se

detecta invasión en nódulos linfáticos) y 2 (4 o más nódulos invadidos). En el caso de la metástasis sólo hay dos posibilidades: 0 si no ha habido metástasis y 1 si sí la hay.

4.2 Alteraciones genéticas en el cáncer colorrectal

Es importante descubrir marcadores tempranos de cáncer colorrectal ya que el diagnóstico precoz de la neoplasia es determinante a la hora de erradicarla. Estudios previos sobre el desarrollo y progresión del cáncer colorrectal identifican una gran batería de genes implicados en el proceso; el evento desencadenante del proceso se desconoce pero se sabe que las células pertenecientes a la neoplasia muestran una serie de alteraciones generalizadas en el genoma como hipometilación del DNA, mutaciones en el gen K-ras y/o pérdida de genes supresores de tumores, debido generalmente a deleciones en el cromosoma 17p (Lengauer *et al*, 1998). Las mutaciones en el gen K-ras aparecen en estadio temprano del desarrollo tumoral aproximadamente un 30% de los tumores de colon (Slattery *et al*, 2001). Este gen forma parte de las vías de señalización celular RAS-MAPK y RAS-PI(3)K, que juegan un papel central en la tumorigénesis a través de la fosforilación y activación de factores de transcripción directamente relacionados con el crecimiento, diferenciación y apoptosis celular (Shaw y Cantley, 2006).

La pérdida de parte del cromosoma 17 se produce en más del 75% de los carcinomas colorrectales; la región comúnmente perdida de este cromosoma contiene el gen P53. P53 es un gen supresor de tumores al que se ha llamado guardián del genoma por su función en la regulación del ciclo celular. La proteína p53 induce el arresto del ciclo celular o la apoptosis en respuesta a un daño severo en el ADN regulando la transcripción de genes que detienen el crecimiento celular e inhiben el crecimiento e invasión celulares.

El 85% de los casos de cáncer colorrectal presenta mutaciones en el gen supresor de tumores *APC*, responsable de la poliposis adenomatosa familiar. La proteína APC regula la ruta de señalización celular WNT, una de las vías moduladoras de la proliferación celular. Se denomina Wnt a un grupo de glicoproteínas secretadas que actúan como ligandos en la activación de vías de transducción de señales mediadas por receptores. Las vías Wnt modulan la proliferación celular en el intestino (van de Wetering *et al*, 2002). La vía Wnt/ β -catenina es la más estudiada; sus componentes presentan un alto grado de conservación a lo largo de la evolución de las especies

aunque aparecen proteínas específicas de cada tejido. En ausencia de señalización mediada por Wnt, un complejo de degradación citoplásmico, entre cuyos componentes se encuentra la proteína APC y la axina, regula la fosforilación de la β -catenina mediante la glicógeno sintetasa kinasa 3 (GSK-3), promoviendo la interacción de la β -catenina con la proteína β -TrCP (*b-transducin-repeat-containing protein*) dando lugar a la ubiquitinización de la β -catenina y su degradación en el proteosoma (Polakis, 1999). La activación de la ruta de señalización Wnt mediante la interacción de las proteínas Wnt con receptores serpentina Frizzled (Bhanot *et al*, 1996) y correceptores como LRP5 y LRP6 (*Low-density lipoprotein receptor related protein*) (Pison *et al*, 2000) deriva en la reducción de la actividad de la GSK-3 y, como consecuencia, se reducen los niveles de fosforilación y degradación de la β -catenina, que se acumula en el núcleo. Con el aumento de los niveles de β -catenina en el núcleo se activa la maquinaria de transcripción de sus genes diana mediada por Tcf-4 (*T cell-specific transcription factor 4*) (Korinek *et al*, 1997). Se ha observado una activación constitutiva de la ruta Wnt/ β -catenina en células portadoras de una mutación inactivadora del gen APC, con la consiguiente activación de la transcripción de genes que promueven la proliferación celular, entre ellos c-Myc, un oncogén (Alitalo *et al*, 1983).

Otro de los eventos genéticos altamente relacionado con la neoplasia colorrectal es la inestabilidad de microsatélites. El 15% de los tumores esporádicos, así como los casos de cáncer hereditario no-polipósico, presenta deficiencias en el sistema de reparación del DNA (Woerner *et al*, 2003), responsable de la aparición de inserciones y deleciones a lo largo del genoma que se hacen muy evidentes en los microsatélites dando lugar a la inestabilidad de microsatélites. Los microsatélites son una parte de los cromosomas sin función conocida formada por secuencias de bases cortas repetitivas que por su estructura están predispuestas a acumular errores durante la replicación del DNA. Estos errores son reparados por los complejos proteicos codificados por los genes reparadores (*hMLH1*, *hMLH2*, *hMLH6*, *hMSH3*, *hPMS1* y *hPMS2*). La disfunción de los genes reparadores, que se detecta analizando la estabilidad de los microsatélites (Ionov *et al*, 1993), genera un estado de hipermutabilidad en todos los genes de la célula, entre ellos los genes reguladores del ciclo celular, proto-oncogenes y genes supresores de tumores. Los tumores que presentan inestabilidad de microsatélites

presentan una secuencia adenoma-carcinoma acelerada. Sin embargo, presentan un menor grado de metástasis por lo que tienen mejor pronóstico (Søreide *et al*, 2006).

5. Marcadores tumorales

Los marcadores tumorales son moléculas producidas por las células tumorales u otras células del tejido neoplásico que permiten discernir entre un individuo sano y uno enfermo.

Alteraciones en las secuencias y niveles de expresión génica y proteica han sido asociadas a diferentes tipos de cáncer. Estos marcadores moleculares pueden ser muy útiles a la hora de detectar el cáncer, determinar la pronóstico del mismo y monitorizar su progresión o respuesta a los tratamientos.

Actualmente, el desarrollo de las nuevas tecnologías “ómicas” que permiten analizar un gran número de genes y proteínas simultáneamente ha hecho posible el estudio a nivel global de las características diferenciales de un tumor y ha dado lugar a un gran avance en el descubrimiento de nuevos marcadores asociados al cáncer. Los métodos tradicionales para la identificación de los genes implicados en la progresión tumoral se basan en el estudio de genes individuales. En la última década, el desarrollo de la tecnología de los *microarrays* de DNA (Shalon *et al*, 1996) ha permitido a los investigadores pasar del estudio de un gen individual al estudio de muchos genes simultáneamente. Comparando los perfiles de expresión de una muestra normal frente a una tumoral se pueden definir patrones de expresión génica propios de un determinado tipo de cáncer. Utilizando los *microarrays* de DNA se han caracterizado varios tipos de cáncer como el de mama (Sorlie *et al*, 2001), próstata (Dhanasekaran *et al*, 2001), vejiga (Thykjaer *et al*, 2001) o melanoma (DeRisi *et al*, 1996) entre otros, permitiendo caracterizar los genes que se expresan de forma anómala en dichos tumores.

Aunque en la mayoría de los casos la carcinogénesis tiene una base genética, es la alteración de los productos proteicos la que modifica el fenotipo de una célula convirtiéndola en tumoral. Muchas modificaciones postraduccionales de las proteínas, algunas de ellas asociadas al desarrollo del cáncer, no se pueden determinar a partir de la información genética de los genes implicados. La proteómica ha dado lugar al descubrimiento de nuevos marcadores tumorales; al igual que la genómica, analiza simultáneamente un gran número de proteínas en células o tejidos tumorales y sanos.

Entre los métodos utilizados en la proteómica se encuentra el 2D-DIGE (*two-dimensional differential in gel electrophoresis*) que se basa en el marcaje diferencial de las proteínas de dos muestras distintas (tumoral vs normal) y la separación de las mismas según su carga y tamaño (Marouga *et al*, 2005). Comparando los patrones de expresión del tejido normal con los del tumoral se han identificado proteínas cuya abundancia o modificación se relaciona con el desarrollo y progresión del cáncer (Alfonso *et al*, 2005). La espectrometría de masas también se ha utilizado para comparar perfiles de expresión entre tejidos sanos y tumorales; tecnologías como la espectrometría de masas MALDI-TOF (*matrix-assisted laser-desorption-ionization time-of-flight*) o LC-ESI-MS/MS (*liquid-chromatography-ion-spray tandem mass spectrometry*) utilizan la fragmentación de las proteínas presentes en las muestras tras digerirlas con enzimas proteolíticas y la separación de los fragmentos mediante técnicas cromatográficas. Para cada muestra se genera un espectro de fragmentación que permite diferenciar muestras normales de tumorales según los fragmentos proteicos generados (Cañas *et al*, 2006).

La proteómica también se utiliza para detectar las proteínas y/o auto-anticuerpos presentes en el suero de pacientes con cáncer. Las células tumorales secretan al suero diferentes proteínas que pueden ser utilizadas como marcadores tumorales; el estudio de los niveles de estas proteínas en suero, o sus anticuerpos correspondientes, puede ser eficaz a la hora de hacer un diagnóstico temprano del cáncer. Los *microarrays* de proteínas se han convertido en un método muy útil a la hora de analizar a gran escala las diferencias en los niveles de expresión proteicos entre tejidos o sueros sanos y tumorales (Madoz-Gúrpide *et al*, 2007). Las proteínas que presentan modificaciones o alteraciones en su nivel de expresión en la progresión tumoral constituyen una buena diana tanto para el diagnóstico como para el tratamiento del cáncer.

Actualmente, son pocos los marcadores proteicos disponibles para diagnosticar o monitorizar el cáncer. Por ejemplo, el PSA (*prostate-specific antigen*) (Stamey *et al*, 1987) se utiliza en el diagnóstico del cáncer de próstata. Sin embargo, niveles altos de esta proteína no siempre están ligados a la presencia de un tumor, generando un gran número de falsos positivos. Igualmente, los niveles en suero de CEA (*carcinoembryonic antigen*) (Gold & Freedman, 1975) se utilizan para diagnosticar el cáncer colorrectal, sin embargo, no todos los pacientes con cáncer colorrectal presentan niveles alterados

de esta proteína, dando lugar a falsos negativos. Otro de los marcadores más utilizado es el CA-125 (Bast *et al*, 1981) para diagnosticar el cáncer de ovario pero, al igual que el PSA en cáncer de próstata, da lugar a un gran número de falsos positivos ya que los niveles aumentados de esta proteína en suero no siempre están relacionados con un proceso neoplásico.

5.1 Marcadores en cáncer colorrectal

Desde un punto de vista molecular, el cáncer colorrectal es uno de los tumores sólidos más estudiados. Sin embargo, se conocen pocos marcadores de cáncer que representan potenciales dianas terapéuticas. Para definir nuevos marcadores tumorales en este tipo de neoplasia es necesario llevar a cabo estudios que permitan traducir a un nivel proteómico los avances generados a nivel genómico. En los últimos años se ha enfatizado en el estudio de la localización de las proteínas en las células y tejidos y de los niveles de expresión diferenciales de dichas proteínas entre tejidos sanos y tumorales. Estos estudios se basan en la producción de anticuerpos específicos frente al producto proteico de cada gen implicado en el desarrollo y progresión tumorales para determinar no sólo su localización celular sino también su abundancia.

Por otro lado, el desarrollo de la proteómica ha permitido identificar directamente proteínas cuya abundancia está involucrada en la progresión tumoral, independientemente de los niveles de expresión de los genes que las codifican. La aplicación de técnicas proteómicas permite determinar las modificaciones postraduccionales presentes en las proteínas de las células y los tejidos, pudiendo ser alguna de dichas modificaciones determinantes a la hora de otorgar un fenotipo tumoral a la célula.

6. Sistemas de expresión de proteínas

Para validar a nivel proteómico el conocimiento alcanzado a nivel genómico sobre el cáncer colorrectal es necesario disponer de un método rápido de producción de las proteínas derivadas de los genes identificados. Las proteínas obtenidas pueden utilizarse tanto para desarrollar anticuerpos específicos que permitan analizar a nivel histológico los niveles de expresión proteica de los genes sobre-expresados o inhibidos en cáncer y su localización celular y tisular, como para realizar *microarrays* de proteínas que permitan analizar los autoanticuerpos presentes en el suero de los pacientes.

Los organismos más utilizados como sistemas de expresión de proteínas han sido, sobre todo, *E. coli*, células de insecto (baculovirus) y levaduras, como *Pichia pastoris*. Cada uno de estos sistemas de expresión presenta una serie de ventajas e inconvenientes teniendo en cuenta el coste económico, la facilidad de manejo y la necesidad o no de obtener las proteínas con modificaciones postraduccionales.

El grupo del Dr. LaBaer realizó un estudio sobre la facilidad de obtener proteínas humanas a gran escala en *E. coli* evitando la necesidad de centrarse en las características particulares de cada una de ellas (Braun *et al*, 2002). Este trabajo confirma la versatilidad de *E. coli* para la obtención de un gran número de proteínas simultáneamente y la influencia del *tag* en la expresión y solubilidad de las proteínas. Otro estudio aborda la expresión de diferentes proteínas humanas de forma simultánea utilizando como sistema de expresión las células de insecto, mediada por la infección con baculovirus (Albala *et al*, 2000), señalando como principal ventaja de este sistema de expresión sobre *E. coli* la presencia de modificaciones postraduccionales en las proteínas exógenas obtenidas dado que se trata de células eucariotas.

El primer paso en la producción *in vitro* de una proteína es la construcción de un vector de expresión adecuado. Existen diferentes *tags* o marcas que se pueden fusionar a la proteína diana, presentando cada uno de ellos una serie de ventajas e inconvenientes a valorar dependiendo del uso posterior que se quiera dar a la proteína de interés.

Los diferentes *tags* disponibles influyen en la estabilidad, solubilidad y niveles de expresión de las proteínas a las que se fusionan en *E. coli*; así, la MBP (Proteína de unión a la maltosa) y la tiorredoxina (Trx) ayudan a solubilizar las proteínas en *E. coli* (Hammarström *et al*, 2002) mientras que el pequeño tamaño de la cola de polihistidinas hace de este *tag* uno de lo más utilizados dado que no afecta al plegamiento de las proteínas a las que se fusiona.

6.1 Expresión de proteínas en *E. coli*

Escherichia coli es el vehículo de expresión de proteínas recombinantes más utilizado. Es un sistema muy fácil de manipular, de bajo coste y rápido. Sin embargo, se trata de un organismo procariota por lo que no se introducen modificaciones postraduccionales en las proteínas expresadas. Entre las ventajas de este sistema de expresión cabe destacar la rapidez con la que se obtienen las proteínas y su bajo coste. Por otro lado, una vez transformadas las células bacterianas se obtienen clones estables

que se conservan durante años, pudiendo escalar los cultivos de expresión a grandes volúmenes. Otra de las ventajas del uso de este sistema es la capacidad de regular la expresión de las proteínas exógenas; para evitar la citotoxicidad de las proteínas introducidas los vectores de expresión presentan diferentes mecanismos de represión de la expresión.

Entre los inconvenientes que presenta la expresión de proteínas en *E. coli* está la ausencia de modificaciones postraduccionales y el alto porcentaje de proteínas que se obtiene como productos insolubles debido a que *E. coli* tiende a formar cuerpos de inclusión.

6.2 Sistemas de expresión de alto rendimiento. Tecnología Gateway

El uso de técnicas tradicionales de clonación basadas en la digestión y ligación de fragmentos de DNA es uno de los obstáculos en la expresión de proteínas a gran escala porque no son muy eficientes y consumen mucho tiempo. Uno de los modos de superar este inconveniente es la utilización de sistemas de clonación recombinante como *Gateway* (Walhout *et al*, 2000). Este sistema trabaja en dos pasos consecutivos en los que primero se crea un vector de entrada y después múltiples vectores de destino con diferentes *tags* para diferentes sistemas de expresión. Los dos pasos a seguir son muy eficientes y fáciles de realizar con muchos clones a la vez lo que permite obtener un gran número de proteínas simultáneamente.

La tecnología *Gateway* se basa en las reacciones de recombinación que median la integración y escisión del fago λ en el genoma de *E. coli* (Landy, 1989). La integración del fago implica la recombinación del sitio attP del DNA del fago con el sitio attB del genoma bacteriano. Esta integración requiere la actuación de dos enzimas: la Integrasa (Int), del fago y el Factor de Integración del Huésped (IHF), de la bacteria. La reacción de recombinación es reversible de forma que el genoma del fago puede ser extraído del genoma de la bacteria mediante recombinación entre los sitios attL y attR. La reacción de escisión requiere de las enzimas Int, IHF y de una enzima adicional sintetizada por el fago: la Escisionasa (Xis) (Figura 3).

El esquema de clonaje en la tecnología de *Gateway* se muestra en la Figura 4. El producto de PCR de la proteína de interés (cDNA) se introduce por recombinación en un vector de entrada al sistema llamado pDONOR o pENTR (*entry clone*). El vector resultante es utilizado como vector de entrada a los vectores destino de expresión.

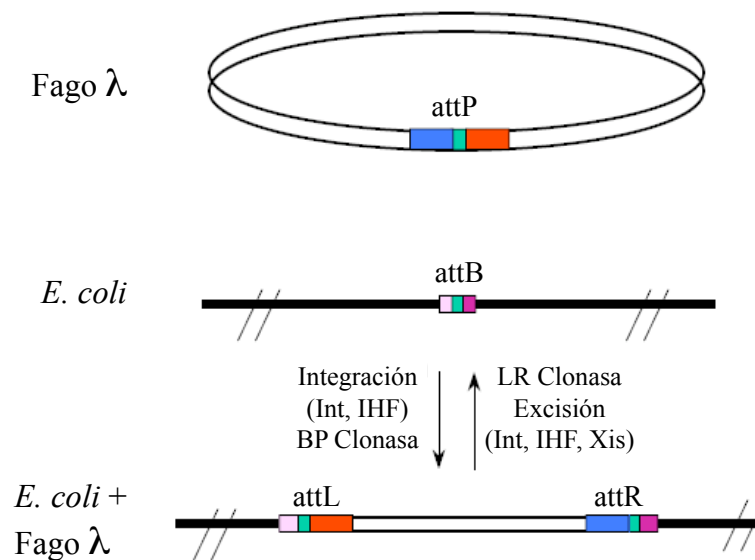


Figura 3. Recombinación entre los sitios $attP$ vírico y $attB$ bacteriano. La tecnología *Gateway* basa su principio de recombinación homóloga para construir los vectores de entrada y de destino en el mecanismo de inserción en el cromosoma bacteriano del fago λ (Invitrogen).

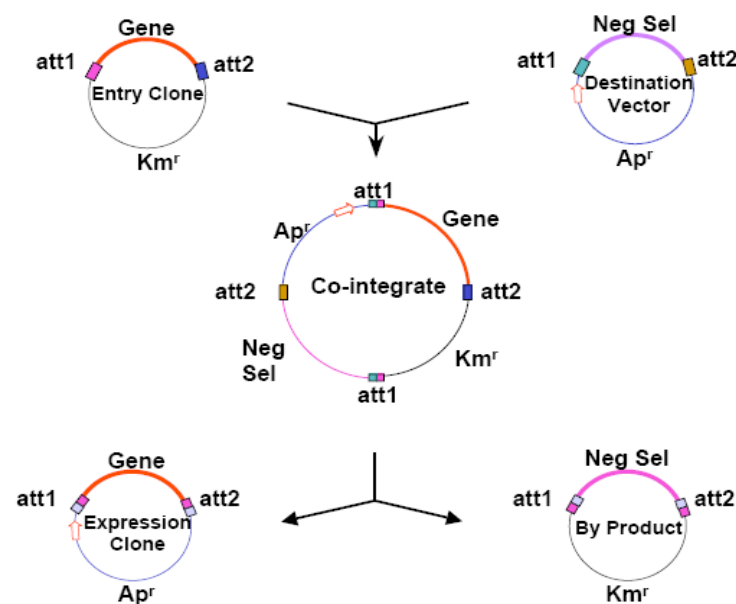


Figura 4. Reacción de recombinación homóloga entre el vector de entrada y el de destino. La duplicación de los sitios att en los vectores permite que la recombinación entre ambos sea direccional. El vector saliente sin inserto es letal para las células portadoras (Invitrogen).

Otro modo de llevar a cabo la clonación direccional en el sistema *Gateway* es mediante la estrategia TOPO. En este caso, la reacción de recombinación se lleva a cabo entre el gen de interés y el vector pENTR/D-TOPO, que contiene la topoisomerasa. El único requisito necesario es añadir la secuencia CACC al oligonucleótido 5' que se vaya a utilizar para amplificar el gen de interés. Como se muestra en la Figura 5, en presencia

de la topoisomerasa, el saliente GTGG del vector invade el extremo 5' del producto de PCR, hibrida con las bases CACC añadidas y estabiliza el producto de PCR en la orientación correcta.

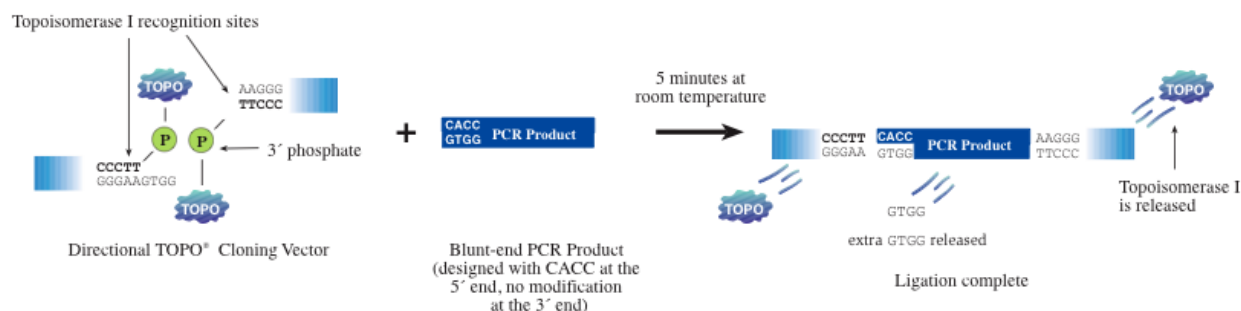


Figura 5. Reacción de recombinación direccional TOPO. La enzima Topoisomerasa I actúa como enzima de restricción y ligasa. La adición de la secuencia CACC al extremo del oligonucleótido 5' empleado en la PCR permite insertar en la dirección correcta el producto de PCR en el vector (Invitrogen).

7. Anticuerpos en la investigación y la terapia del cáncer

Uno de los mayores obstáculos a la hora de confirmar los resultados obtenidos con los análisis genómicos a nivel proteómico es la ausencia de anticuerpos que reconozcan específicamente las proteínas derivadas de los diferentes estudios. En muchos casos no es posible evaluar las dianas obtenidas porque no hay anticuerpos disponibles para confirmar la expresión de las proteínas identificadas en muestras reales y líneas celulares.

Por otro lado, los anticuerpos son la clase de moléculas terapéuticas que más rápidamente está creciendo desde los años 90. Se utilizan en el tratamiento de diferentes enfermedades como el cáncer, inflamaciones crónicas, rechazo de transplantes, enfermedades infecciosas y cardiovasculares.

De forma clásica, los anticuerpos se obtienen a partir de sueros de animales inmunizados frente a un determinado antígeno (anticuerpos policlonales). Los anticuerpos monoclonales (Köhler y Milstein, 1975) se diferencian de los policlonales en que los primeros reconocen un único epítipo del antígeno mientras que los segundos reconocen varios epítopos ya que no existe una especie única de anticuerpo en el suero del animal. En los últimos años se han desarrollado nuevas estrategias para la selección de anticuerpos, entre ellas el despliegue en fagos (McCafferty *et al*, 1990) y la obtención de anticuerpos monoespecíficos (Agaton *et al*, 2003), obtenidos tras

inmunizar a los animales con fragmentos proteicos de secuencia única que representan una proteína completa (PrEST, *Protein Epitope Signature Tag*).

8. Anticuerpos monoclonales

La tecnología desarrollada para la producción de anticuerpos monoclonales fue la de los hibridomas de ratón generados mediante la fusión estable de células de mieloma con células B de ratones inmunizados frente a un determinado antígeno (Köhler y Milstein, 1975). Pese al éxito que tuvo esta técnica en la investigación y en el descubrimiento de nuevos anticuerpos, no tuvo un gran éxito en clínica debido a la alta inmunogenicidad que presentan los anticuerpos de ratón en humanos y a la débil activación del complemento humano. Estas limitaciones de los anticuerpos de ratón se han ido solventando mediante la quimerización o humanización de los mismos. La quimerización de los anticuerpos consiste en la fusión de las regiones variables del anticuerpo de ratón, aquellas que reconocen al antígeno, a las regiones constantes de los anticuerpos humanos (Morrison *et al*, 1984; Boulianne *et al*, 1984). La humanización de los anticuerpos requiere la transferencia de las regiones determinantes de complementariedad (CDRs, *Complementary-determining regions*), que son los *loops* de reconocimiento del antígeno, a la IgG (Inmunoglobulina G) humana (Jones *et al*, 1986; Verhoeven *et al*, 1988; Riechmann *et al*, 1988).

Los anticuerpos utilizados tanto en ensayos clínicos como en investigación presentan una gran gama de formatos. La arquitectura globular de los anticuerpos facilita la creación de diferentes especies moleculares en las que varía tanto el número de regiones constantes presentes como el número de regiones variables de unión al antígeno (Figura 6).

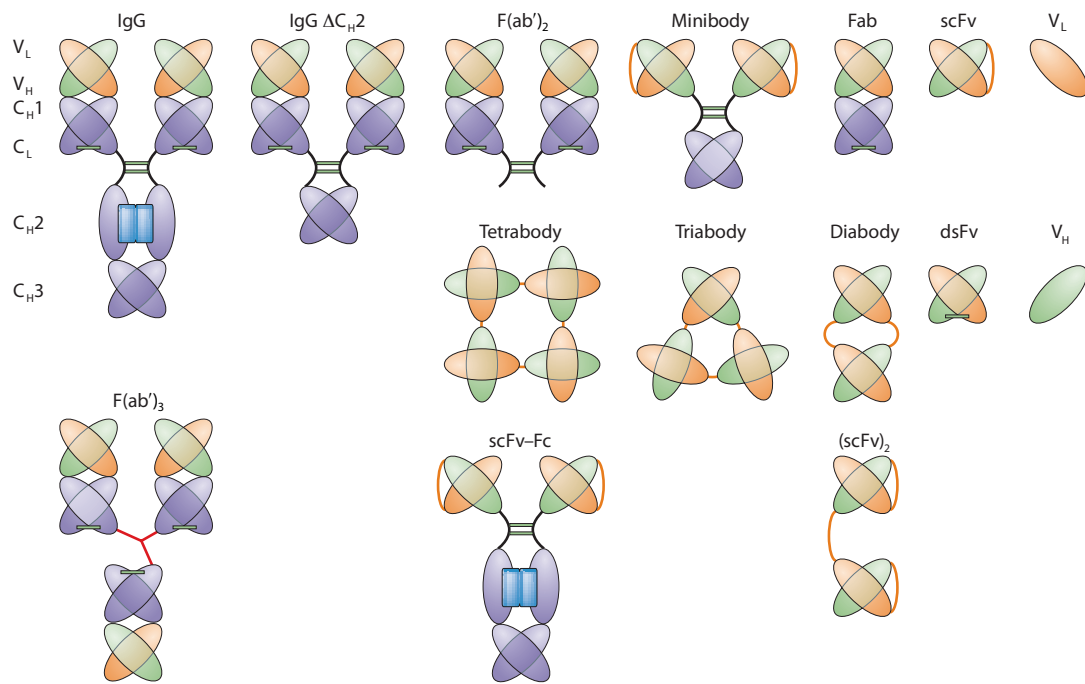


Figura 6. Representación de los diferentes formatos de los anticuerpos. Los anticuerpos nativos tienen una estructura globular que permite generar diferentes formatos de un mismo anticuerpo (Carter, 2006).

9. Tecnología del despliegue en fagos

Los anticuerpos que llegan a la fase clínica deben ser humanos. Se pueden obtener mediante la tecnología del despliegue en fagos (*Phage display*) (McCafferty *et al*, 1990) o mediante la expresión de los genes de las inmunoglobulinas humanas en ratones transgénicos (Lonberg, 2005).

La tecnología del despliegue en fagos fue desarrollada por Smith (Smith, 1985), quien utilizó el bacteriófago filamentoso M13 como sistema de despliegue de péptidos. A pesar de la existencia de varios sistemas de despliegue alternativos, el fago M13 continúa siendo el método más utilizado.

El fago M13 es un fago filamentoso no lítico formado por un genoma de cadena sencilla circular de 6407 bases. La cápside está formada por copias de 5 proteínas siendo la mayoritaria la pVIII, con 2700 copias situadas a lo largo del fago (Barbas, 2001); en cada extremo se sitúan dos proteínas, con 5 copias de cada una de ellas: pVII y pIX en un extremo y pIII y pVI en el otro (Figura 7A).

Este fago infecta bacterias de *E. coli* portadoras del pili F, codificado por el plásmido conjugativo F. Las partículas víricas se unen al pili F y el genoma del fago se transloca en el citoplasma bacteriano. El genoma del fago se replica utilizando tanto

proteínas bacterianas como víricas y se empaqueta en el interior de las células infectadas en partículas que son secretadas al medio. Las partículas víricas tienen que ser transportadas al periplasma bacteriano antes de ser ensambladas y expulsadas al exterior, completándose así el ciclo celular (Figura 7B).

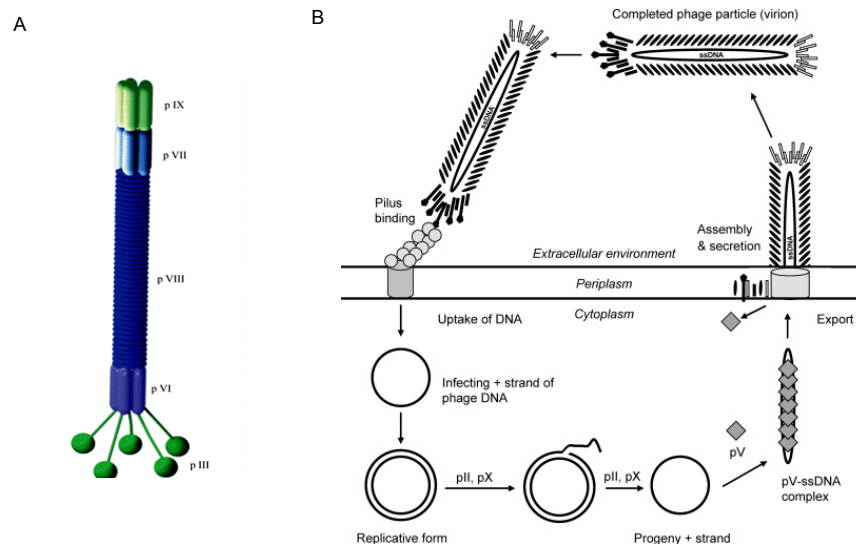


Figura 7. Esquema de la morfología y ciclo vital del bacteriófago M13. A) Morfología del virus bacteriófago M13. La proteína de la cápside más utilizada para fusionar péptidos es la pIII. B) Ciclo vital del fago M13: los fagos se unen a *E. coli* a través del pili. Tras la unión al fago el pili se retrae y acerca al fago a la superficie celular. Una vez allí, el extremo N-terminal de la proteína pIII interacciona con la proteína de la membrana bacteriana TolA. El DNA de cadena sencilla (ss+) del fago entra en el interior de la célula y se convierte en DNA de doble cadena (forma replicativa) mediante proteínas propias de *E. coli*. Dos proteínas virales, pII y pX, están involucradas en la amplificación del DNA de doble cadena y la producción de múltiples copias de ss+ DNA. Estas copias del DNA se unen a la proteína pV y el complejo interacciona con las proteínas de exportación pI, pIV, pIX y tiorredoxina, que ensamblan y secretan las partículas virales. Las partículas virales abandonan la célula sin destruirla (Kehoe y Kay, 2005).

Inicialmente se empleaban fagos que contenían toda la información necesaria para completar el ciclo vital del fago pero finalmente se ha optado por utilizar los fagémidos como vectores de despliegue. Los fagémidos son pequeños vectores con una gran eficiencia de transformación que los hace muy atractivos para construir librerías de gran repertorio; portan el gen III, que codifica la proteína pIII con sitios apropiados para el clonaje de los genes foráneos en el extremo N-terminal de la proteína, y el origen de replicación del fago. Las proteínas necesarias para completar el ciclo vital del fago las aporta un fago cooperador, generalmente M13KO7 o VCSM13, que suple al fago filamentoso de todas las proteínas estructurales y poseen un origen de replicación defetivo. Dado que el fago cooperador codifica para la proteína pIII *wild-type*, generalmente el 90% de los fagos producidos carecen de la proteína de fusión en la pIII;

este problema se ha solucionado utilizando fagos cooperadores con el gen III deleccionado (Griffiths *et al*, 1993; Duenas y Borrebaeck, 1995)

Mediante sucesivas rondas de selección, se seleccionan péptidos o proteínas, incluidos anticuerpos, de gran afinidad y especificidad frente a prácticamente cualquier diana. Una de las principales ventajas de esta técnica es la relación directa que existe entre el genotipo y el fenotipo del fago, permitiendo el análisis y mejora de las proteínas expuestas para modificar sus propiedades. En el caso de los anticuerpos la posibilidad de manipular el DNA que los sintetiza permite manipular su tamaño, valencia, afinidad y fusionarlos a diferentes efectores (sustancias tóxicas para las células, proteínas fluorescentes, etc.).

10. Librerías de anticuerpos desplegados en fagos

La tecnología del despliegue en fagos permite exponer librerías de péptidos o proteínas en la superficie del fago M13. Los anticuerpos fueron las primeras proteínas desplegadas en la superficie del fago en formato de *single-chain Fv* (scFv) (McCafferty *et al*, 1990); se fusionaron las secuencias codificadoras de las regiones variables tanto de la cadena pesada como de la ligera al extremo N-terminal de la proteína pIII de la cápside vírica. Las inmunoglobulinas presentan una disposición globular de los dominios por lo que las modificaciones de la molécula en zonas de no-reconocimiento del antígeno no alteran la capacidad de unión del anticuerpo al antígeno. La secuencia nucleotídica de la scFv se clonó en fase con el gen III bajo la influencia de la secuencia pelB, que dirige la secreción de la proteína pIII al periplasma. En el ambiente reductor del periplasma las scFv se pliegan correctamente mediante la formación de puentes disulfuro intramoleculares. La presencia de la scFv fusionada a la proteína pIII no influye en la infectividad de las partículas víricas.

Existen diferentes tipos de librerías que pueden ser utilizadas a la hora de seleccionar anticuerpos frente a un determinado antígeno: librerías inmunes, *naïve* y sintéticas.

1. Librerías de anticuerpos de animales inmunizados o donantes inmunes. El repertorio de anticuerpos se obtiene a partir de los genes de las inmunoglobulinas G de las células B del bazo de animales inmunizados con el antígeno o de donantes inmunes. De esta manera, el repertorio de anticuerpos está enriquecido en anticuerpos específicos

frente al antígeno de interés. Teniendo en cuenta que las células B productoras de anticuerpos del bazo o del plasma son accesibles, este tipo de librerías son muy útiles a la hora de analizar la respuesta humoral de pacientes con enfermedades autoinmunes, infecciones virales o procesos neoplásicos.

2. Librerías *naïve* de donantes no inmunizados. En este tipo de librerías, el repertorio de anticuerpos está formado por un conjunto de inmunoglobulinas M que reconocen una gran variedad de antígenos. Este conjunto de anticuerpos puede ser clonado como un repertorio *naïve* de genes reordenados tras amplificar los genes variables a partir del mRNA de las IgM de las células B.

3. Librerías sintéticas. En este tipo de librerías, los anticuerpos se manipulan para introducir en las Regiones Determinantes de Complementariedad (CDR) un nivel predeterminado de variabilidad, sobre todo en el CDR3 de la cadena pesada, ya que naturalmente este *loop* presenta la mayor diversidad.

En esta tesis doctoral se utilizaron las librerías Tomlinson I y J y las librerías Mehta. Las librerías Tomlinson son librerías de anticuerpos humanos recombinantes creadas en el laboratorio de Greg Winter en el Laboratorio de Biología Molecular del MRC (Cambridge). Las dos librerías se basan en la estructura de los anticuerpos humanos uniendo las regiones variables de las cadenas pesada y ligera *in vitro* y aumentando su diversidad por mutación de las CDRs mediante PCR y oligonucleótidos aleatorios. La librería I tiene un tamaño de $1,47 \times 10^8$ y la librería J de $1,27 \times 10^8$ fagos.

Las librerías Mehta son librerías *naïve* de anticuerpos recombinantes humanos creadas por el grupo de Wayne Marasco en el Instituto Dana-Farber (Boston). Tienen un tamaño de $1,5 \times 10^{10}$ fagos, considerablemente mayor al de las librerías Tomlinson.

La tecnología del despliegue en fagos ha permitido obtener numerosos anticuerpos humanos de interés terapéutico. Las scFv presentan la ventaja de que debido a su pequeño tamaño tienen mayor penetración en el tumor que los anticuerpos convencionales. Esta característica es de gran utilidad a la hora de localizar un tumor mediante técnicas de imagen *in vivo*, fusionadas a radioisótopos (Begent *et al*, 1996), y para desarrollar terapias más específicas frente a las células tumorales fusionando los anticuerpos seleccionados frente a un determinado marcador tumoral a sustancias citotóxicas y dirigir así el tratamiento únicamente a las células portadoras de dicho

marcador, reduciendo considerablemente los efectos secundarios de la quimioterapia en los pacientes.

11. Estudios de afinidad. Uso de Biacore

Uno de los principales problemas que presenta la técnica del despliegue en fagos para la obtención de anticuerpos específicos es la baja afinidad de los anticuerpos seleccionados por el antígeno. Uno de los métodos más utilizados a la hora de valorar las constantes cinéticas en la unión de dos moléculas es la resonancia del plasmón en superficie (SPR, *Surface Plasmon Resonance*) (Aslan *et al*, 2005), un fenómeno óptico que permite la detección a tiempo real de la interacción entre dos analitos no marcados. Esta técnica permite determinar las constantes cinéticas de asociación (K_{on}) y disociación (K_{off}) entre dos moléculas así como la constante de afinidad entre ellas (K_D) sin ningún tipo de marcaje de las moléculas en estudio, evitando por tanto cualquier modificación que el marcaje pueda ejercer en el reconocimiento y la unión de las dos moléculas.

12. *Protein Epitope Signatura tag (PrEST)*

Para facilitar la obtención de anticuerpos a gran escala se ha propuesto un nuevo modelo de expresión de proteínas en *E. coli* que evita las dificultades de expresar determinadas proteínas recombinantes completas en este sistema. Este modelo se basa en la expresión de fragmentos proteicos llamados PrEST (*Protein Epitope Signatura tag*) (Agaton *et al*, 2003). Estos fragmentos proteicos se diseñaron para contener epítomos únicos, válidos para generar anticuerpos. Los fragmentos se seleccionan por ordenador buscando secuencias únicas que no pertenezcan a regiones transmembrana ni a péptidos señal para evitar la insolubilidad en *E. coli*. Para evitar las reacciones cruzadas de los anticuerpos, las regiones y dominios homólogos a otras proteínas humanas se descartan tras utilizar programas de búsqueda de homologías como el BLAST (Altschul *et al*, 1997). Los fragmentos proteicos seleccionados tienen un tamaño suficientemente pequeño como para ser producidos en *E. coli* sin dificultad pero suficientemente grandes como para contener epítomos conformacionales. Los fragmentos expresados y purificados se utilizan para inmunizar animales. Una vez obtenidos los sueros, los anticuerpos monoespecíficos se purifican mediante

cromatografía de afinidad utilizando el propio fragmento proteico seleccionado como ligando de la columna.

La obtención de los anticuerpos de una manera rápida gracias al desarrollo de los *PrEST* permite conocer mejor la expresión de las proteínas, su función y su localización. El proyecto del Atlas de Proteínas Humanas pretende elaborar el perfil de expresión de todas las proteínas humanas tanto en tejidos sanos como portadores de diferentes patologías mediante técnicas inmunohistoquímicas utilizando anticuerpos específicos para cada proteína (Uhlén *et al*, 2005). Este nuevo método de obtención de fragmentos representativos de las diferentes proteínas del proteoma humano podría facilitar en gran medida la disponibilidad de los anticuerpos específicos necesarios para determinar su localización y niveles de expresión.

13. Validación de dianas genómicas a nivel proteómico

Este trabajo de tesis doctoral estudia la validación a nivel proteómico de las dianas genómicas descubiertas a partir de un análisis de expresión realizado con muestras tumorales y normales de pacientes con cáncer colorrectal, utilizando microarrays de cDNA (Oncochip CNIO) (Tracey *et al*, 2002) para identificar genes diferencialmente expresados entre mucosa tumoral y normal.

A partir de la lista completa de genes alterados se seleccionó un panel de genes correspondientes a proteínas de membrana y extracelulares y se expresaron las proteínas correspondientes en *E. coli*, estas proteínas se utilizaron para la producción de anticuerpos específicos utilizando para ello tanto las técnicas clásicas (anticuerpos poli y monoclonales) como las nuevas (despliegue en fagos y anticuerpos monoespecíficos).

Una vez obtenida la batería de anticuerpos frente a las diferentes dianas seleccionadas, se estudió mediante inmunohistoquímica y *western blot* el valor de los marcadores genómicos a nivel proteómico analizando la expresión proteica diferencial entre tejidos sanos y tumorales.

Finalmente, se estudió la influencia de tres genes en la proliferación y la apoptosis de líneas celulares de cáncer de colon; para ello, se utilizaron las técnicas de silenciamiento de genes mediante RNA de interferencia y transfección de células.

Objetivos

Durante el desarrollo de esta tesis doctoral se plantearon los siguientes objetivos:

1. Identificación de nuevos marcadores en cáncer colorrectal por análisis transcriptómico y posterior validación proteómica de los genes alterados.
 - 1.1. Análisis del transcriptoma en el cáncer colorrectal utilizando oncochips CNIO para el estudio de la expresión génica diferencial en 22 muestras pareadas tumoral/normal.
 - 1.2. Estudio de la ontología y localización celular de los genes alterados. Selección de proteínas de membrana y extracelulares.
 - 1.3. Validación de dianas por PCR cuantitativa tanto en muestras tumorales como en líneas celulares de cáncer de colon.
2. Estudio y valoración de métodos para la expresión rápida y simultánea en *E. coli* de múltiples proteínas correspondientes a las dianas genómicas seleccionadas.
3. Análisis de diferentes métodos para la producción rápida de anticuerpos y validación de dianas.
 - 3.1. Uso de técnicas combinatoriales como despliegue en fagos.
 - 3.2. Uno de la tecnología PrEST.
 - 3.3. Validación mediante técnicas inmunológicas de las diferencias en los niveles de expresión proteica entre tejidos sanos y tumorales de cáncer colorrectal.
4. Estudios funcionales de las dianas seleccionadas en la proliferación y la apoptosis. Análisis del efecto de la sobreexpresión y silenciamiento de las dianas seleccionadas en diferentes líneas de cáncer colorrectal.

Materiales y Métodos

1. Cultivo de líneas celulares

Las líneas celulares de cáncer colorrectal CaCo2, Colo205, Co115, Hct15, Hct116, RKO, Sw48 y Sw480 se crecieron en medio RPMI (Sigma) suplementado con suero fetal bovino (Gibco) al 10% y penicilina y estreptomycin (Gibco) al 1% en un incubador humidificado a 37° C de temperatura y un 5% de CO₂.

2. Uso de *microarrays* de cDNA en muestras de cáncer colorrectal

Se reunieron 22 muestras de tejido de pacientes con cáncer colorrectal obtenidas a través del banco de tejidos del CNIO en colaboración con diferentes hospitales españoles: Virgen de la Salud (Toledo), Cartagena (Murcia), Puerta de Hierro (Madrid), Clínico de San Carlos (Madrid), Ramón y Cajal (Madrid) y Alcorcón (Madrid). De cada paciente se extrajo una muestra de tejido tumoral y de tejido normal adyacente. La historia clínica de cada paciente se muestra en la Tabla 1.

Tabla 1. Datos clínicos de las muestras incluidas en el análisis de la expresión génica

Nº Biopsia	Nº Tumor	Nº Normal	Edad (años)	Sexo (a)	Localización del tumor (b)	Estadio Dukes
1010118	36T	35N	78	H	CA	C
1010283	80T	78N	78	H	CA	C
1010286	92T	71N	45	M	CD	C
1010292	81T	72N	63	H	CD	C
1010320	77T	90N	71	H	CA	C
1010322	96T	95N	72	M	CD	C
1010414	107T	106N	81	M	CD	C
1010433	84T	87N	56	M	CA	C
1010445	83T	85N	79	M	CA+CD	C
1010455	109T	108N	67	H	CD	C
1010457	111T	110N	81	M	CA	C
3011002	1T	15N	68	H	CD	B
3011003	59T	37N		H	CA	B
3011011	20T	26N	80	M	CA	A
3011021	30T	29N	40	M	CA	B
5011390	62T	41N	75	H	CA	A
5011392	48T	47N	71	M	CD	B
6010532	40T	39N	87	M	CA	A
6010533	46T	45N	54	M	CA	C
9010070	54T	53N	77	H	CA	C
9010085	56T	55N	80	H	CA	B
9010086	58T	57N	78	M	CD	C

H: Hombre; M: Mujer

CA: Colon ascendente; CD: Colon descendente

El RNA total se aisló a partir de 50 o 100 mg de las biopsias congeladas según los protocolos descritos anteriormente (Tracey *et al*, 2002; Martínez *et al*, 2003). Para generar cDNA fluorescente se marcaron 30 µg de RNA tumoral y sano con Cy5-dUTP y Cy3-dUTP (GE Healthcare) respectivamente.

Se utilizaron los *microarrays* CNIO (CNIO Oncochip) para hibridar las muestras. Cada *microarray* hibridado se escaneó con un *scanner* Agilent G2565BA y las imágenes se cuantificaron con la aplicación GenePix Pro 4.0 (Axon Instruments Inc.). Tras la normalización de los datos con la rutina “Lowess” (Lin *et al*, 2002; Smyth y Speed, 2003), se realizaron test múltiples para detectar genes diferencialmente expresados utilizando el método de análisis significativo de *microarrays* (SAM) (Tusher *et al*, 2001).

3. PCR Cuantitativa en tumores y líneas celulares

3.1 Extracción de RNA total y síntesis de cDNA a partir de tejido y líneas celulares

En el caso de los tejidos, la extracción del RNA total se llevó a cabo añadiendo 1, 5 ml de Trizol por cada 100 mg de tejido y homogenizando con un homogenizador Ultra-Turrax T8 (IKA), a continuación se incubó la muestra durante 5 min a temperatura ambiente. En el caso de las células, se centrifugaron durante 5 minutos a 1000 rpm tras levantarlas de la placa de cultivo incubando 5 minutos con tripsina-EDTA al 0,5% (Gibco) y se resuspendió el precipitado en 1 ml de Trizol (Invitrogen) por cada 5×10^6 células, a continuación se incubó durante 5 minutos a temperatura ambiente. En ambos casos, se añadieron 200 μ l de cloroformo y se incubaron durante 3 minutos a temperatura ambiente. Se centrifugó la mezcla a 11.000 rpm durante 15 minutos a 4° C, se añadieron 500 μ l de isopropanol a la fase acuosa y se incubó durante 10 minutos a temperatura ambiente. Se centrifugó la mezcla durante 10 minutos a 10.000 rpm a 4° C y se descartó el sobrenadante. Se añadieron 700 μ l de etanol absoluto (Merck) al precipitado que contiene el RNA y se purificó con Rneasy kit (Qiagen) según las instrucciones del fabricante. La concentración del RNA total obtenido se midió en un espectrofotómetro Nanodrop ND-1000 (Nanodrop Technologies) y su calidad se analizó en un BioAnalyzer (Applied Biosystems).

La síntesis del cDNA se llevó a cabo por RT-PCR a partir de 3 μ g de RNA total de cada línea celular con la enzima Superscript III RNasa H Reverse Transcriptasa (Invitrogen) empleando oligo dT como cebador. Brevemente, en una primera reacción se mezclaron 3 μ g de RNA, 1 μ l de oligo dT, 1 μ l de dNTPs y agua DEPC en un

volumen final de 13 μ l y se incubaron durante 5 min a 65° C y durante 2 min a 4° C. A continuación, se añadió a cada tubo de reacción 4 μ l de tampón First-strand, 1 μ l de DTT 0,1 M, 1 μ l de RNasa OUT y 1 μ l de Superscript III RNasa H RT y se incubó a 25° C durante 5 min, a 60° C durante 60 min y a 70° C durante 15 min. Por último, se añadió 1 μ l de RNasa H y se incubó durante 20 min a 37° C. Las concentraciones de cDNA obtenidas se midieron en el Nanodrop (Applied Biosystems).

3.2 PCR Cuantitativa

La PCR cuantitativa se realizó utilizando como cDNA de partida 100 ng de cDNA de cada línea. Como control negativo se utilizó cDNA obtenido a partir de tejido de colon sano. Las PCR se llevaron a cabo en un volumen final de 10 μ l con oligonucleótidos específicos de cada gen marcados con FAM (6-carboxi-fluorescein) (Applied Biosystems) a una concentración final 0,25 μ M y el tampón universal de PCR TaqMan (Applied Biosystems) que incluye dNTPs, DNA polimerasa, tampón de la polimerasa y MgCl₂. Como control interno se amplificó el gen GUSB. La reacción de amplificación se llevó a cabo en placas de 384 pocillos en un termociclador ABI 7900HT (Applied Biosystems). En primer lugar se calentaron las muestras a 50° C durante 2 min y a 95° C durante 10 min, a continuación se llevaron a cabo 50 ciclos de 95° C durante 1,5 min y 60° C a 1 min. Cada muestra y gen se amplificó por triplicado.

Para el análisis de la PCR cuantitativa se utilizó el programa SDS 2.2.2 (Applied Biosystems) y el estudio $\Delta\Delta C_t$, siendo C_t el ciclo a partir del cual la muestra alcanza una cantidad determinada de cDNA dentro de la fase de amplificación exponencial; este valor lo determina el programa al establecer un C_t automático y elimina a su vez los datos considerados anómalos. Cada resultado se normalizó frente a los niveles de GUSB y se comparó la cantidad de cDNA de cada gen en cada línea o muestra tumoral con la cantidad en las muestras normales.

4. Expresión de proteínas

4.1 Predicción de regiones transmembrana y péptidos señal

Para facilitar la expresión de proteínas solubles en *E. coli* se predijo la presencia de regiones transmembrana y péptidos señal de las proteínas seleccionadas. Para estudiar la presencia de regiones transmembrana se utilizó el algoritmo SOSUI (Hirokawa *et al*,

1998) (www.sosui.proteome.bio.tuat.ac.jp). Este algoritmo permitió predecir estas regiones y decidir si se eliminaban o no, evitando así la obtención de proteínas insolubles. Para la identificación de péptidos señal se utilizó el programa SMART (Schultz *et al*, 1998) (www.smart.embl-heidelberg.de).

4.2 Obtención de los clones de cDNA

Se seleccionaron 29 genes para generar sus correspondientes productos proteicos. Se utilizaron tres fuentes para obtener los clones de partida. Clones IMAGE (*Integrated Molecular Analysis of Genome Expression Consortium*) (Research Genetics) (Lennon *et al*, 1996) con cDNA de secuencias parciales de la fase de lectura abierta (ORF). El requisito que se estableció para utilizar clones de esta librería fue que el cDNA contuviese al menos el 30% del ORF total. Alternativamente se utilizó la colección MGC (*Mammalian Gene Collection*) (MRC Geneservices) que contiene clones de cDNA de secuencia completa. Por último, los clones para SULF1 y THBS2 fueron proporcionados por el Dr. Rosen (Universidad de California) y Dr. Bernstein (Universidad de Washington) respectivamente.

Todos los clones utilizados fueron secuenciados para confirmar su secuencia utilizando un secuenciador de DNA automático (ABI7002, Applied Biosystems).

4.3 Diseño de oligonucleótidos

Para amplificar las regiones codificadoras de los clones se diseñaron oligonucleótidos de forma que se eliminasen los péptidos señal para facilitar la expresión soluble en *E. coli*. Para un clonaje rápido de los genes en el sistema *Gateway* se utilizó el vector pENTR/D-TOPO (Invitrogen). Los oligonucleótidos directos se diseñaron con la secuencia TOPO (CACC) delante del codon inicial ATG, dando un tamaño medio de 28 nucleótidos. Los oligonucleótidos reversos se diseñaron de forma que introdujesen un nuevo codon *stop* en los clones con secuencias parciales. Todos los oligonucleótidos se solicitaron a Sigma Genosis. En las Tablas 2 y 3 se muestran los oligonucleótidos utilizados con los clones de la colección IMAGE y MGC, respectivamente.

Table 2. Oligonucleótidos utilizados par la amplificación por PCR de los genes seleccionados a partir de la colección IMAGE.

Gen	Oligonucleótido	Secuencia
CA9	Directo	CACCGCACTCCTGCCCTCTGACTTCAGC
	Reverso	CTAGGCTCCAGTCTCGGCTACCTCTGC
THBS2	Directo	CACCGGTCACCAGGACAAAGACACGACC
	Reverso	GAGCATGAAGACCAAGGTGACCAGTGG
BMP4	Directo	CACCGCGAGCCATGCTAGTTTGATACCT
	Reverso	CTATTCCCGGGTCCAGCGAAGGACCGC
SPP1	Directo	CACCGAGACCCTTCCAAGTAAGTCCAACGAA
	Reverso	CTAATCAATCACATCGGAATGCTC
MMP7	Directo	CACCGTTGCAGAATACTCACTATTTC
	Reverso	CTATTTCTTTCTTGAATTACTTCT
CTHRC1	Directo	CACCAGTGCTCTAAGAGTTTGTTCAGT
	Reverso	TTATTTTGGTAGTTCTTCAATAATGAT
MMP11	Directo	CACCGGTGAAAAGCCAGTCCTGGGCCCC
	Reverso	TCAGAGGAAAAGTGTTGGCAGGCTCGG
LCN2	Directo	CACCGCCCAGGACTCCACCTCAGACCTG
	Reverso	CTAGAAGTACTCCCTGTTTTGAGA
IFITM3	Directo	CACCGTCCAAACCTTCTTCTCTCCTGTC
	Reverso	CTATCCATAGGCCTGGAAGATCAGCAC
IFITM2	Directo	CACCATGAACCACATTGTGCAAACCTTC
	Reverso	CTATCGCTGGGCCTGGACGACCAACAC
IFITM1	Directo	CACCATGCACAAGGAGGAACATGGGGTG
	Reverso	CTAGTACACCCGTTTTTCCTGTAT

Table 3. Oligonucleótidos utilizados par la amplificación por PCR de los genes seleccionados a partir de la colección MGC.

Gen	Oligonucleótido	Secuencia
PTK7	Directo	CACCGCCATTGTCTTCATCAAGCAGCCG
	Reverso	TCACGGCTTGCTGTCCACGGTGCTGTC
EPHB3	Directo	CACCGGCTGCCGGGCGCTGGAAGAGACC
	Reverso	TCAGACCTGCACAGGCAGCGTCTG
SULF1	Directo	CACCTCGACTGTCAGATCCCCGAGGTTCA
	Reverso	TTAACCTTCCCATCCATCCATAACTGT
SLC26A3	Directo	CACCATGATTGAACCCTTTGGAATCAG
	Reverso	TTAGAAATTTGTTTCAACTGGCACCTC
AGT	Directo	CACCACCGGGTGACATACACCCCTTC
	Reverso	TCATGCTGTGCTCAGCGGGTTGGCCAC
CA9	Directo	CACCCAGAGGTTGCCCGGATGCAGGAG
	Reverso	TCATGCTGTGCTCAGCGGGTTGGC
PLAU	Directo	CACCAGCAATGAACCTCATCAAGTTCCA
	Reverso	TCAGAGGGCCAGGCCATTCTCTCTCTT
THBS2	Directo	CACCAGGACAAAGACACGACC
	Reverso	AGAGCATGAAGACCAAGGTGACC
BMP4	Directo	CACCAGCCCTAAGCATCACTCACAGCGG
	Reverso	TCAGCGGCACCCACATCCCTCTAC
SEPP1	Directo	CACCGAGAGCCAGGACCAAAGCTCCTTA
	Reverso	TTAGTTTGAAGGTCATTCTCACTTTT
ANXA3	Directo	CACCATGGCATCTATCTGGGTTGGACAC
	Reverso	TCAGTCATCTCCACCACAGATTT
SPP1	Directo	CACCATGAGAATTGCAGTGATTGCTTT
	Reverso	TTAATTGACCTCAGAAGATGCACTATC
SPARC	Directo	CACCATGAGGGCCTGGATCTTCTTTCTC
	Reverso	TTAGATCACAAGATCCTTGTCGATATC
IGFBP3	Directo	CACCGGCGCGAGCTCGGGGGGCTTGGGT
	Reverso	CTACTTGCTCTGCATGCTGTAGCAGTG
MMP7	Directo	CACCCTGCCGCTGCCTCAGGAGGCGGGA
	Reverso	CTATTCTTTCTTGAATTACTTCTCTT
CTHRC1	Directo	CACCGAGATCCCCAAGGGGAAGCAAAAG
	Reverso	TTATTTTGGTAGTTCTTCAATAAT
MMP11	Directo	CACCTTCGTGCTTTCTGGCGGCGCTGG
	Reverso	TCAGAGGAAAGTGTGGCAGGCTCGGC
TIMP1	Directo	CACCATGGCCCCCTTTGAGCCCCTGGCT
	Reverso	TCAGGCTATCTGGGACCGCAGGGACTG
LCN2	Directo	CACCCAGGACTCCACCTCAGACCTGATC
	Reverso	TCAGCCGTCGATACACTGGTCGATTGG
TDGF1	Directo	CACCATGGACTGCAGGAAGATGGCCCGC
	Reverso	TCAGGCTATCTGGGACCGCAGGGACTG
IFITM3	Directo	CACCATGAATCACACTGTCCAAACCTTC
	Reverso	CTATCCATAGGCCTGGAAGATCAG
IFITM2	Directo	CACCATGAACCACATTGTGCAAAACCTTC
	Reverso	CTATCGCTGGGCTGGACGACCAACAC
IFITM1	Directo	CACCGTCTACCATATTATGTTACAGATA
	Reverso	CTAGTAACCCCGTTTTTCTGTAT
CXCL3	Directo	CACCGCGTCCGTGGTCACTGAACTGCGC
	Reverso	TCAGTTGGTGCTCCCTTGTTTCAG
CXCL1	Directo	CACCGCGTCCGTGGCCACTGAACTGCGC
	Reverso	TCAGTTGGATTGTCACTGTTTCAG
HIG2	Directo	CACCATCTTCGTTAGAGTGATGGAGTCC
	Reverso	TCACATGCTTCTGGATGGATGGATGGTC

4.4 Clonaje en el sistema pENTR/D-TOPO

Las regiones codificantes fueron amplificadas a partir de los cDNA mediante PCR con la enzima Vent DNA polimerasa (New England Biolabs), esta enzima tiene actividad exonucleasa correctora 3'-5'. La reacción de PCR se llevó a cabo en un volumen de 50 µl conteniendo oligonucleótidos a una concentración final de 0,5 µM, dNTP (Roche) 200 µM, tampón de amplificación 1x (New England Biolabs), BSA y 1U de polimerasa. Como programa de amplificación se utilizó 1 ciclo de 3 min a 94° C, 25 ciclos (1 min a 94°C, 1 min a 55° C y 2 min a 72 °C) y un ciclo final de 7 min a 72° C.

A continuación, 1 µl del producto de PCR se mezcló con 1 µl del vector TOPO, se dejó 5 minutos a temperatura ambiente y se transformaron células *E. coli* TOP10 competentes (Invitrogen), siguiendo las instrucciones recomendadas por la casa comercial.

La eficiencia de la transformación se analizó mediante PCR de colonia tras hacer una placa réplica. Las colonias se lisaron con agua para liberar el DNA molde para la reacción. Las PCR se realizaron en un volumen final de 10 µl con los oligonucleótidos T7 directo 5'TAATACGACTCACTATAGG 3', y reverso (específico de cada cDNA) a una concentración final de 0,5 µM, tampón 1x, MgCl₂ 1,5 mM, dNTP 0,2 mM (Roche) y 1,5 U de Taq.

A partir de la placa réplica se inoculó una colonia positiva en 10 ml de LB suplementado con 50 µg/ml de kanamicina y se purificaron los vectores por miniprep utilizando el kit Wizard (Promega), siguiendo las instrucciones del fabricante. Los plásmidos se secuenciaron en un secuenciador automático (ABI7002, Applied Biosystems) antes de generar los vectores de destino (pDEST).

5. Clonaje en los vectores destino

Se utilizaron tres vectores destino diferentes: pDEST15 (Invitrogen), que contiene GST (Glutation S Transferasa), pDEST17 (Invitrogen), que contiene 6xHis (cola de 6 histidinas) y pTH1 (donado por el profesor T. Hard del *Royal Institute of Technology*, Estocolmo, Suecia), que contiene MBP (*Maltose Binding Protein*, proteína de unión a la maltosa).

Para generar los vectores destino se mezclaron cantidades equivalentes del vector pENTR/D-TOPO y del vector destino (300 ng), se añadió 1 µl de la enzima LR Clonasa (Invitrogen) y se incubó durante 1h a 25° C. Para inactivar la reacción de la clonasa se añadieron 2 µl de Proteinasa K (Invitrogen) y se incubó la mezcla durante 10 min a 37° C. Células competentes BL21 (DE3) (Edge Biosystems) fueron transformadas directamente con el producto de la clonación mediante choque térmico incubando 30 µl de células con 2 µl de vector durante 30 min en hielo. A continuación se incubaron a 42° C durante 45 s y se pasaron a hielo durante 2 min. Se añadieron 900 µl de medio SOC (Tryptona al 20% (p/v), extracto de levadura al 5% (p/v), NaCl 8,55 M, KCl 0,25 mM, MgCl₂ 2 M y Glucosa 1 M) y se incubaron a 37° C con agitación durante 45 min. Se sembraron 100 µl y el precipitado celular en placas de LB-Agar (Tryptona al 2% (p/v), extracto de levadura al 1% (p/v), NaCl 1,17 M y Agar bacteriológico al 1,5% (p/v)) suplementadas con 100 µg/ml de ampicilina y se dejaron crecer durante toda la noche a 37° C.

En aquellos casos en que la transformación directa en células BL21 (DE3) no fue efectiva, se transformaron células DH5α (Invitrogen), de mayor eficiencia, para luego purificar los vectores y transformar BL21 (DE3). Se utilizó el mismo protocolo en ambos casos.

6. Expresión de proteínas en *E. coli*

Colonias individuales de células BL21 (DE3) transformadas con los vectores destino en medio LB suplementado con 100 µg/ml de ampicilina se crecieron toda la noche a 37° C con agitación. Se relanzaron los cultivos después de una dilución 1/100 en 1L de medio LB conteniendo 100 µg/ml ampicilina y se crecieron a 37° C con agitación hasta alcanzar la fase de crecimiento exponencial (DO₆₀₀ 0,6-0,8), momento en el que se añadió isopropil 1-β-D-galactopiranosido (IPTG) a una concentración final de 0,4 mM. Tras 3 horas de inducción, las células se centrifugaron a 4000 rpm y el precipitado se resuspendió en 0,05 volúmenes de PBS suplementado con inhibidores de proteasas (Complete, Roche). Para lisar las células, se llevaron a cabo tres ciclos de congelación/descongelación y a continuación se dieron 2 pulsos de sonicación de 5 s a una amplitud del 10% (Digital sonifier, Branson). Tras la lisis celular, las muestras se

centrifugaron a 10.000 rpm durante 10 minutos para separar la fracción soluble de la insoluble.

7. Electroforesis en geles de poliacrilamida y *western blot*

Las proteínas diluídas en tampón Laemmli (1970) (SDS al 2%, 2-mercaptoetanol 280 mM, azul de bromofenol al 0,1%, Tris-HCl pH 6,8 100 mM, glicerol al 10%) y desnaturalizadas durante 5 min a 100° C se resolvieron mediante electroforesis en minigeles (10 cm x 10 cm x 0,75 mm) de poliacrilamida-SDS al 10%. En primer lugar se preparó el gel separador con acrilamida-bisacrilamida (29:1) (Bio-Rad) al 10%, Tris-HCl 0,38 M pH 8,8, SDS 0,1 %, persulfato amónico 0,1% y TEMED (tetrametiletilendiamida) 0,01% y se dejó polimerizar a temperatura ambiente tras añadir 1 ml de isopropanol en la parte superior para evitar la inhibición por oxígeno. Finalmente, se retiró el isopropanol y sobre el gel separador se cargó el gel concentrador formado por acrilamida-bisacrilamida (29:1), Tris-HCl pH 6,8 0,125 M, SDS 6,5%, PSA 0,1% y TEMED 0,01%. Se colocaron los peines de 0,75 mm de espesor y se dejó polimerizar a temperatura ambiente.

Las condiciones de electroforesis fueron 220 V (sin limitar amperaje) en tampón de electroforesis (Trizma base 25 mM, Glicina 190 mM, SDS 8,5 mM) utilizando una fuente Power-pac/300 (Bio-Rad). La tinción de los geles se llevó a cabo con azul brillante de *Coomassie* G250 (Bio-Rad).

Para el *western blot* las proteínas separadas por electroforesis se transfirieron a membranas de nitrocelulosa Hybond-C (Amersham Biosciences) utilizando el equipo de transferencia semiseca Trans-Blot SD de (Bio-Rad) durante 45 min a 22 V (sin limitar amperaje) en tampón de transferencia (Trizma base 25 mM, Glicina 190 mM, Metanol al 20%). Las membranas se bloquearon con PBS suplementado con Tween-20 (Sigma) al 0,05% y leche semidesnatada en polvo al 3% durante 1h a temperatura ambiente. A continuación, se incubaron las membranas con anticuerpos específicos frente a cada *tag* (Monoclonal anti MBP (Sigma) a una dilución 1:10.000, scFv anti GST (propio del laboratorio), anti His (Roche) a una dilución 1:5.000 durante 2h a temperatura ambiente. A continuación se incubaron con el anticuerpo secundario correspondiente conjugado a peroxidasa o fosfatasa alcalina durante 1h a temperatura ambiente. La visualización se realizó con ECL (Amersham Biosciences) para anticuerpos conjugados a peroxidasa o

NBT/BCIP (tetrazolio/5-bromo-cloro-3-indolil-fosfato) (Bio-Rad) para aquellos conjugados a fosfatasa alcalina, en este último caso, la solución reveladora está compuesta por Tris-HCl 1 M pH 9,5, MgCl₂ 0,5 mM, NBT a una dilución 1:100 y BCIP a una dilución 1:100.

8. Purificación de las proteínas recombinantes

8.1 Purificación de proteínas fusionadas a MBP

Las proteínas solubles fusionadas a MBP se purificaron mediante cromatografía de afinidad. Se utilizaron 5 ml de la fracción soluble y 1 ml de resina de amilosa (New England Biolabs) equilibrada con Tris HCl 20 mM pH 7,4, NaCl 200 mM, EDTA 1 mM. Para la elución se utilizaron 10 ml del tampón de equilibrado suplementado con maltosa 10 mM. La purificación de las proteínas a maltosa se llevó a cabo en columnas (Bio-Rad) cargadas con la resina de amilosa. Se lavó la resina con 10 volúmenes de tampón de equilibrado dejando fluir la solución por gravedad. Se introdujo la muestra en la columna y se dejó fluir a lo largo de la resina, a continuación, se lavó la resina con 20 volúmenes de resina de tampón de equilibrado, se eluyeron las proteínas unidas a la resina con 10 volúmenes de tampón de elución y se recogieron en fracciones de 1 ml. Las fracciones obtenidas se analizaron por electroforesis en geles de poliacrilamida-SDS y posterior tinción con Azul brillante de coomassie.

8.2 Purificación de proteínas fusionadas a GST

Para la purificación de proteínas solubles fusionadas a GST se utilizaron columnas GSTrapTM (Amersham Biosciences) de 1 ml acopladas a un equipo ÄKTA Prime (Amersham Biosciences). El tampón de equilibrado de la columna contiene fosfato sódico 20 mM pH 7,3, NaCl 0,15 mM. Se purificaron 5 ml de fracción soluble. El tampón utilizado para la elución está compuesto por Tris HCl 50 mM pH 8,0, glutation reducido 10 mM. Todos los tampones fueron filtrados (Sartorius, 0,45mm) y degasificados. Brevemente, el programa de purificación de proteínas fusionadas a GST utilizado consta de un paso de equilibrado de la columna con 10 ml de tampón de equilibrado a un flujo de 5 ml/min, inyección de la muestra en la columna a un flujo de 1 ml/min (5 ml), lavado de la columna para eliminar las proteínas unidas inespecíficamente con 10 ml de tampón de equilibrado y elución de las proteínas recombinantes con 10 ml de tampón de elución a un flujo de 1 ml/min. Las proteínas

recombinantes eluyen, generalmente, en las fracciones 2, 3 y 4, dichas fracciones se dializaron frente a PBS en membranas de diálisis (Medicell Internacional Ltd) durante toda la noche con agitación para eliminar el glutatión.

8.3 Purificación de proteínas fusionadas a 6xHis

La purificación de las proteínas fusionadas al *tag* de histidinas se llevó a cabo mediante cromatografía de afinidad a metal inmovilizado (IMAC) utilizando columnas de Ni^{2+} HiTrap™ de 1 ml acopladas a un sistema ÄKTA Prime (Amersham Biosciences). Para cargar de Ni^{2+} la columna se utilizó NiSO_4 100 mM. Como tampón de equilibrado de la columna se utilizó fosfato sódico 20 mM pH 7,4, NaCl 0,5 M, imidazol 20 mM (Merck), y como tampón de elución el mismo de equilibrado con una concentración de imidazol 0,5 M. Todos los tampones fueron filtrados (Sartorius, 0,45 μm) y desgasificados. Brevemente, el programa de purificación de proteínas fusionadas a 6xHis consta de los siguientes pasos: lavado de la columna con 25 ml de agua y tampón de equilibrado de la columna a un flujo de 40 ml/min, aplicación de 1 ml de NiSO_4 a un flujo de 1 ml/min, equilibrado de la columna con 10 ml de tampón de equilibrado a un flujo de 1 ml/min, inyección de 5 ml de muestra a un flujo de 1 ml/min, lavado de la columna con 25 ml de tampón de equilibrado y elución de las proteínas unidas a la columna con un gradiente de imidazol desde 20 mM (tampón de equilibrado) hasta 0,5 M (tampón de elución) mezclando diferentes porcentajes de cada solución. Las fracciones recogidas en las que estaban presentes las proteínas recombinantes se dializaron frente a PBS en membranas de diálisis durante toda la noche a 4° C con agitación para eliminar el imidazol.

8.4 Replegamiento de las proteínas insolubles

En el caso de las proteínas insolubles, las proteínas fusionadas a 6xHis se purificaron mediante renaturalización en columna, utilizando como tampón de solubilización urea 6M, Tris HCl 20 mM pH 8,0, NaCl 0,5 M, imidazol 5 mM, 2-mercaptoetanol 1 mM. Como tampón de equilibrado de la columna se utilizó cloruro de guanidinio 6 M en Tris HCl 20 mM pH 8,0, NaCl 0,5 M, imidazol 5 mM, 2-mercaptoetanol 1 mM y como tampón de elución Tris HCl 20 mM pH 8,0, NaCl 0,5 M, imidazol 0,5 M, 2-mercaptoetanol 1 mM. Como tampón de renaturalización de las proteínas se utilizó el mismo tampón que para el equilibrado de la columna pero sin cloruro de guanidinio. Se inyecta el tampón de renaturalización gradualmente, haciendo

un gradiente del tampón desde 0 hasta 100% a un flujo de 0,5 ml/min. A continuación se eluyen las proteínas unidas a la columna y renaturalizadas con el tampón de elución a un flujo de 1 ml/min, disminuyendo paulatinamente la concentración de tampón de renaturalización desde 100 hasta 0%.

9. Producción de anticuerpos recombinantes mediante despliegue en fagos

9.1 Obtención de anticuerpos a partir de las librerías Tomlinson

9.1.1 Preparación de las librerías

Las librerías Tomlinson I y J (MRC Geneservices) se inocularon por separado en 200 ml de medio 2xTY (16 g/l de triptona, 10 g/l de extracto de levadura y 5 g/l de NaCl) suplementado con 100 µg/ml de ampicilina y un 1% de glucosa. Se crecieron a 37° C en agitación hasta alcanzar un DO₆₀₀ de 0,4. Se añadieron 2x10¹¹ fagos KM13 (MRC Geneservices) (Kristensen y Winter, 1998) a 50 ml de cada cultivo y se incubaron durante 30 min a 37° C. Se centrifugó cada cultivo a 3.000g durante 10 min y se resuspendió el precipitado en 100 ml de 2xTY suplementado con 100 µg/ml de ampicilina, 50 µg/ml de kanamicina y 0,1% de glucosa. Se incubó durante toda la noche a 30° C con agitación.

Para amplificar los fagos, se centrifugaron los cultivos a 3.300g durante 30 min, se añadieron 20 ml de PEG/NaCl (Polietilenglicol 6000 (Sigma) al 20%, NaCl 2,5 M) a 80 ml de sobrenadante, se mezcló bien y se dejó en hielo durante 1h. Se centrifugaron a 3.300g durante 30 min, se retiró el PEG/NaCl y se resuspendieron los precipitados en 4 ml de PBS. Se centrifugaron de nuevo a 11.600g durante 10 min para eliminar restos de bacteria, se añadió glicerol a una concentración final de 15% y se guardaron las librerías a -80° C.

9.1.2 Titulación del fago

Para la titulación de fagos se inocularon células de *E. coli* TG1 en 2xTY y se cultivaron a 37° C en agitación hasta alcanzar una DO_{600nm} de 0,4. Se prepararon 6 diluciones seriadas 1/100 del fago, se añadió 1µl de cada una de ellas a 200µl de células y se incubaron durante 30 min a 37° C. Se añadieron 3ml de NZY Top agar (5g/l de NaCl, 2g/l de MgSO₄·7H₂O, 5g/l de extracto de levadura, 10g/l de triptona, 15g/l de agar y 0,7% (p/v) de agarosa) a cada incubación y se plaquearon en placas Petri. Tras

incubar durante toda la noche 37° C se contaron las unidades formadoras de calva de cada dilución para obtener el título del fago.

9.1.3 Selección de scFv

Para la selección de scFv específicas a partir de las librerías Tomlinson (MRC Geneservices) se utilizaron como antígenos las proteínas de secuencia completa fusionadas a MBP: CXCL3, MMP7 y SPARC. Se realizaron tres rondas de selección de fagos utilizando placas de 96 pocillos Maxisorp (Nunc). En un primer paso se mezclaron en suspensión las librerías I+J (en una proporción 1/1) con 50 µg de MBP en PBS suplementado con leche semidesnatada en polvo al 2% (p/v). Tras 1h de incubación a 37° C se añadió la mezcla a los pocillos con 1 µg de antígeno inmovilizado y se incubó durante 2h a temperatura ambiente. Se lavaron los pocillos 10 veces (20 veces en las siguientes rondas de selección) con PBS suplementado con Tween 20 (Sigma) al 0.1% (v/v) y se añadieron 100 µl de tripsina (Sigma) para eluir los fagos unidos al antígeno. Se infectaron células de *E. coli* de la cepa TG1 (MRC Geneservices) en crecimiento exponencial (DO_{600} 0,4) durante 30 min a 37° C con los fagos eluidos. Se sembraron las bacterias infectadas en placas de TYE suplementadas con 100 µg/ml de ampicilina y glucosa al 1% (v/v) y se dejaron crecer durante la noche a 37° C. A continuación se recogieron las bacterias de las placas con medio 2xTY suplementado con glicerol al 15% y se diluyeron 50 µl de las bacterias recogidas en 50 ml de medio de cultivo 2xTY suplementado con 100 µg/ml de ampicilina y glucosa al 1%; se incubaron los cultivos hasta alcanzar la fase exponencial y se infectaron las células con 2.5×10^{11} fagos *helper* KM13 (MRC Geneservices). Tras 30 minutos de incubación a 37° C, se centrifugaron los cultivos, se resuspendieron los precipitados en el mismo volumen de medio 2xTY suplementado con 100 µg/ml de ampicilina, 50 µg/ml de kanamicina y glucosa al 0.1% y se crecieron durante toda la noche a 30° C.

Los fagos de cada ronda de selección se precipitaron a partir del medio de cultivo. Para ello se centrifugaron los cultivos a 3.300 g durante 15 minutos y se añadieron 10 ml de PEG/NaCl a 40 ml de sobrenadante y se incubaron 1h en hielo. Se centrifugaron a 3.300 g durante 30 minutos y se descartaron los sobrenadantes. Los precipitados se resuspendieron en 2 ml de PBS y se centrifugaron a 11.600 g durante 10 minutos. Se recuperaron los sobrenadantes y los fagos se utilizaron para la siguiente

ronda de selección.

9.1.4 ELISA de fagos policlonales

Se recubrieron los pocillos de placas de 96 pocillos flexibles (Falcon, BD Biosciences) con 0.3 µg de antígeno en PBS y se dejó durante toda la noche a 4° C. Tras tres lavados con PBS se bloquearon los pocillos con PBS-leche en polvo desnatada al 2% durante 2h a temperatura ambiente. Los fagos se añadieron a los pocillos haciendo diluciones seriadas de factor 1/10 a partir de una dilución inicial 1/10 en PBS Tween 20 al 0,1% y leche en polvo semidesnatada al 2% y se incubaron durante 1 h a temperatura ambiente. Se lavaron los pocillos con PBS-Tween 20 al 0.1% y se incubaron con un anticuerpo anti M13 conjugado a peroxidasa (Amersham Biosciences) a una dilución 1:5.000 durante 1h a temperatura ambiente. Se lavaron de nuevo los pocillos con PBS-Tween 20 al 0.1% y se reveló con 100 µl de TMB (3,3',5 ,5'-tetrametilbenzidina) (Sigma), la reacción colorimétrica se detuvo con 100 µl de ácido sulfúrico 1M y se midió la absorbancia a 450 nm en un lector Model 550 (Bio-Rad).

9.1.5 ELISA de fagos monoclonales

Se picaron colonias individuales de la tercera ronda de selección manualmente o utilizando un picador de colonias automático (Q-pix, Genetix) y se crecieron durante toda la noche a 37° C en agitación en 100 µl de medio 2xTY suplementado con 100 µg/ml de ampicilina y glucosa al 1% en placas de 96 pocillos (Sarstedt). Al día siguiente se diluyeron los cultivos 1/100 en el mismo medio y se incubó durante 2h a 37° C en agitación; a continuación, se añadieron 25 µl de medio con 10⁹ fagos *helper* KM13 y se incubaron durante 1h a 37° C. Se centrifugaron las bacterias y el precipitado celular se resuspendió en 2xTY suplementado con 50 µg/ml de kanamicina y se cultivaron durante toda la noche a 30° C. Finalmente, los cultivos se centrifugaron y se utilizaron 50 µl del sobrenadante para hacer el ELISA en las mismas condiciones que el de fagos policlonales.

9.1.6 ELISA de scFv

Las scFv correspondientes a los clones positivos de secuencia única se expresaron en la cepa supresora de *E.coli* TG1. El procedimiento es básicamente el mismo al descrito para los fagos pero en este caso, cuando se alcanzó la fase exponencial de crecimiento (DO₆₀₀ 0.9), se indujo la expresión de las scFv añadiendo

IPTG a una concentración final de 1 mM y se dejó durante toda la noche a 30° C con agitación. A continuación se centrifugaron los cultivos y se resuspendieron los precipitados celulares en 1/50 volúmenes de tampón TES (Tris-HCl pH 8,0 200 mM, EDTA 0,5 mM, sacarosa 0,5 M) suplementado con 20 µg/ml de benzamidina (Sigma) y 10 µg/ml de inhibidor de tripsina de soja (Sigma) y se diluyó 1,5 veces en tampón 0,2xTES. Se dejaron 30 min en hielo, se centrifugaron a 16.250 g durante 10 minutos y se guardó a 4° C el sobrenadante, correspondiente a la fracción periplásmica. Se utilizaron 50 µl del periplasma para realizar un ELISA frente al antígeno y al *tag*. En este caso se utilizó como anticuerpo secundario anti c-myc conjugado a peroxidasa (Roche) a una dilución 1:2000.

9.1.7 Análisis de las secuencias

Las secuencias codificantes de las scFv positivas por ELISA se amplificaron mediante PCR a partir del vector pIT2 utilizando los oligonucleótidos LMB3 5'-CAGGAAACAGCTATGAC-3' y pHENseq 5'-CTATGCGGCCCCATTCA3' y Taq polimerasa (Figura 8). Los productos de PCR se trataron con las enzimas *ExoI* (USB), para eliminar los oligonucleótidos sobrantes, y *SAP* (Fosfatasa alcalina de gamba) (USB), para eliminar los dNTP sobrantes, a 37° C durante 30 min y a 80° C durante 15 min y se secuenciaron con el oligo LMB3 en un secuenciador automático (ABI7002, Applied Biosystems).

Una vez obtenidas las secuencias, se compararon los clones entre sí para determinar el número de clones de secuencia única obtenidos. Para ello se utilizó la herramienta de traducción del servidor ExPASy Proteomics (<http://www.expasy.ch>) para traducir las secuencias nucleotídicas a aminoácidos y compararlas entre sí con el programa ClustalW.

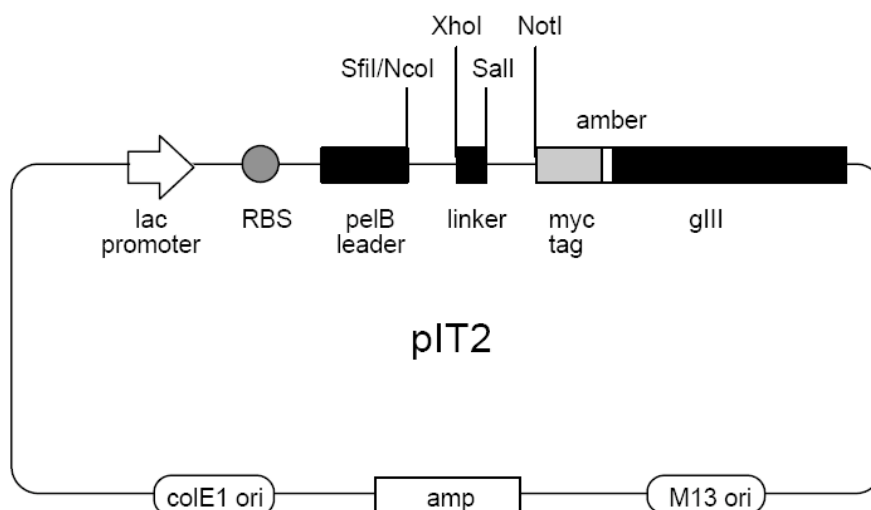


Figura 8. Vector pIT2

9.1.8 Western blot con las scFv seleccionadas

Los *western blot* con las scFv seleccionadas se llevaron a cabo según el procedimiento descrito de electroforesis y transferencia. En este caso, como anticuerpo primario se utilizaron las fracciones periplásmicas a una dilución 1/5 durante 3h a temperatura ambiente. Como anticuerpo secundario se utilizó anti c-myc conjugado a fosfatasa alcalina (Roche) a una dilución 1:2000 y como solución reveladora NBT/BCIP (Bio-Rad).

9.1.9 Eliminación del codon *stop* TAG en las scFv

Los genes de las scFv procedentes de las librerías Tomlinson que contenían codones *stop* internos TAG se amplificaron a partir de los DNA correspondientes mediante PCR utilizando como oligonucleótido directo LMB3 y como oligonucleótido reverso MutAmber (5'-AGACATATG***TT***GC GGCCCCATTCAGATC-3', en negrita y cursiva se señala el nucleótido mutado). El oligonucleótido directo LMB3 es complementario a una secuencia que precede al promotor pelB y el oligo reverso incluye el nuevo codon *stop* TAA, es complementario a la secuencia codificadora del *tag* c-myc y contiene una diana de restricción *NdeI*, presente en el vector pIT2, como sitio de corte único en la secuencia del gen gIII (Figura 9).

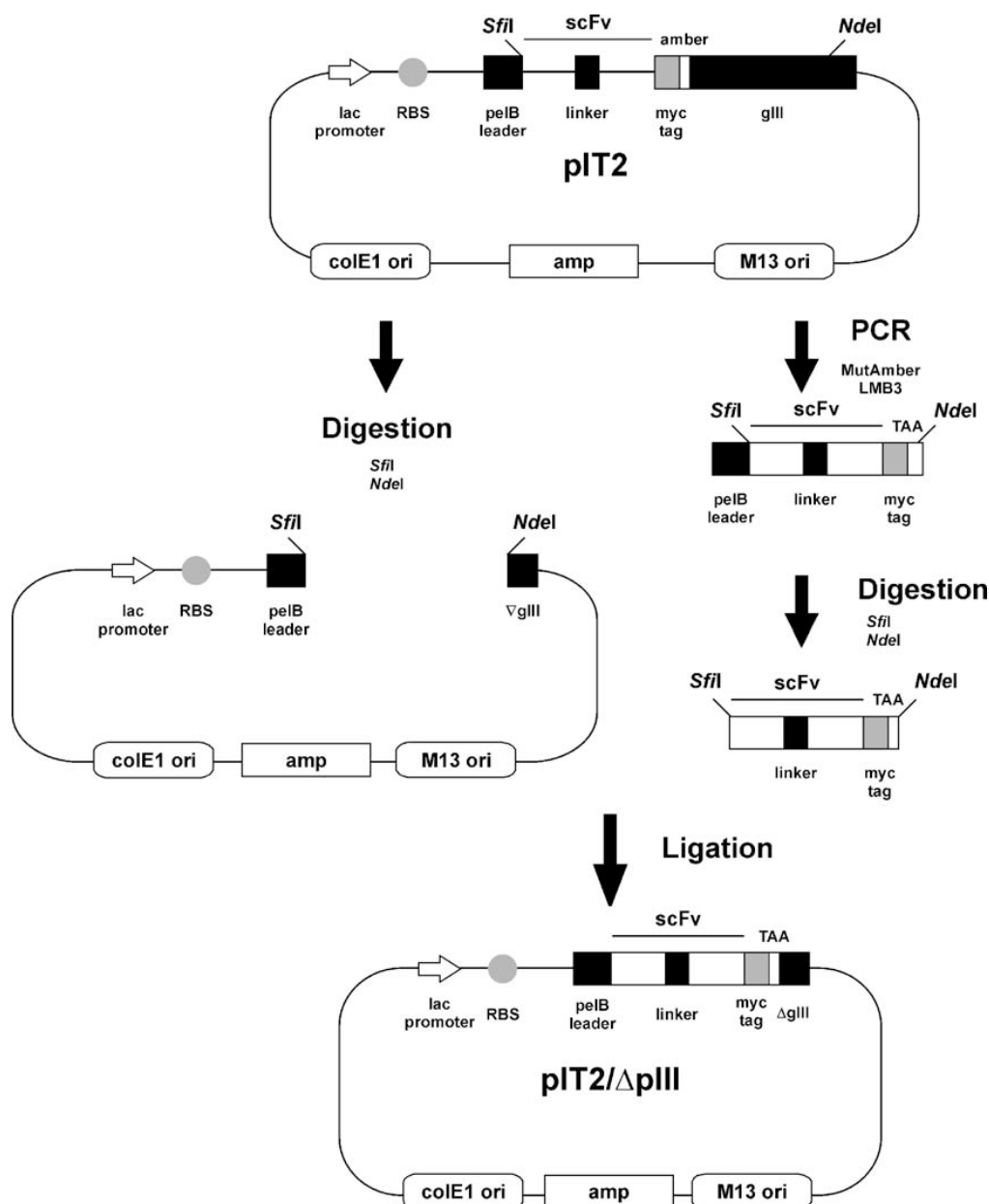


Figura 9. Esquema de la mutagénesis dirigida para la producción de scFv solubles. El codon *stop* TAG presente entre c-myc y el gen gIII es reemplazado por el codon TAA de acuerdo al procedimiento descrito (Barderas *et al*, 2006).

La reacción de amplificación de los genes codificantes de las scFv se llevó a cabo con la enzima KOD DNA polimerasa (Novagen). La reacción de PCR tuvo lugar en un volumen final de 50 μ l con una concentración de cada oligonucleótido de 0,4 μ M, dNTP 200 μ M, tampón de amplificación 1x, MgCl₂ 625 μ M, betaína 1/5 (v/v) y 1U de KOD DNA polimerasa. El programa de amplificación fue 1 ciclo de 3 min a 98° C, 30 ciclos (30 s a 98°C, 30 s a 55° C y 1 min a 72° C) y por último 1 ciclo de 7 min a 72° C.

Los fragmentos amplificados se purificaron con el *kit* QIAquick (Qiagen) siguiendo las instrucciones del proveedor y se digirieron con *SfiI* (Roche) y *NdeI* (New England Biolabs) durante 2h a 37° C en el tampón recomendado por el fabricante. Los fragmentos se subclonaron en el vector pIT2 digerido con las mismas enzimas de restricción utilizando la enzima T4 DNA ligasa (Roche) poniendo una relación molar inserto: vector 3:1 y partiendo de 20 ng de vector por inserto. La ligación se dejó durante una noche a 16° C. Con el producto de la ligación se transformaron células competentes *E. coli* TG1 mediante choque térmico (ver apartado 5).

9.1.10 Expresión y purificación de las scFv sin pIII en TG1

La expresión de las scFv solubles en TG1 y el aislamiento de la fracción periplásmica se llevó a cabo como se explica en el apartado 9.1.6. Se analizaron los niveles de expresión mediante electroforesis en geles de poliacrilamida-SDS al 10% y *western blot* utilizando anti His conjugado a peroxidasa (Sigma) a una dilución 1:1.000 durante 1h a temperatura ambiente.

Las scFv se purificaron mediante cromatografía de afinidad en resina de Co²⁺ (Talon, BD Biosciences). Como tampón de equilibrado de la resina se utilizó fosfato sódico 50 mM pH 7,0, NaCl 300 mM. Se incubaron las fracciones periplásmicas de cada scFv con la resina de Co²⁺ a temperatura ambiente durante 20 min en rotación, tras tres lavados con tampón de equilibrado, se eluyeron las scFv con el tampón de elución (tampón de equilibrado suplementado con imidazol 0,3 M). Las fracciones de elución se analizaron mediante electroforesis en geles de poliacrilamida-SDS al 10 % y tinción con Azul brillante de *Coomassie*.

9.2 Obtención de anticuerpos a partir de las librerías Mehta

9.2.1 Preparación de las librerías Mehta

Las librerías Mehta SJ y AM se amplificaron de la misma manera descrita en la preparación de las librerías Tomlinson I y J (apartado 9.1.1).

9.2.2 Selección de scFv

Para la selección de scFv a partir de las librerías Mehta se utilizaron como antígenos las proteínas de secuencia completa fusionadas a MBP CXCL3, MMP7 y SPARC y la proteína CXCL1 fusionada a GST. La selección de fagos se llevó a cabo en placas de 96 pocillos Maxisorb (Nunc). En la primera ronda de selección se fijó 1 µg de cada proteína, en la segunda ronda se fijaron 0,5 µg de cada proteína y en la tercera 0,1

µg; de esta forma se mejora en cada ronda la afinidad de los fagos seleccionados al disponer cada vez de menor antígeno al que unirse. Los fagos específicos frente a MBP o GST se bloquearon incubándolos previamente con 25 µg del *tag* correspondiente en suspensión. Las 3 rondas de selección se realizaron de la misma manera que con las librerías Tomlinson.

9.2.3 Expresión y purificación de las scFv Mehta en pET28b

Para expresar las scFv seleccionadas a partir de la librería Mehta en el vector pET28b (Novagen) se amplificaron las scFv por PCR a partir del vector pFarber. Las PCR se llevaron a cabo a partir de 15 ng de vector en un volumen final de 50 µl utilizando como oligonucleótido directo pET28-scFvMehtaFw, de secuencia 5' 'CAGTCATCATGAAATACCTATTGCCTAC3' y como oligonucleótido reverso pET28-scFvMehtaRv, de secuencia 5' CATT TTTTCATAATCAAATCACCGGAC TCGAGTGC GGCCCCATT CAGATCCTCT3' a una concentración final de 0,4 µM, dNTPs 0,2 mM, tampón de amplificación 1x, MgCl₂ 625 µM y 1,5 U de KOD DNA polimerasa (Novagen). El programa de amplificación utilizado fue 1 ciclo de 3 min a 98° C, 25 ciclos (15 s a 98° C, 2 s a 60° C, 20 s a 72° C) y un ciclo final de 7 min a 72° C.

Los productos de PCR se purificaron con el kit QIAquick (Quiagen). Se digirieron 2 µg de cada producto de PCR con *RcaI* (Roche) y *XhoI* (Fermentas) durante 2h a 37° C en los tampones recomendados por la casa comercial y 10 µg de vector pET28b con *NcoI* (Fermentas) y *XhoI* (Fermentas) durante 2h a 37° C.

La ligación de insertos y vector se realizó con la enzima T4 Ligasa (Roche) partiendo de 20 ng de vector por inserto. Se transformaron células competentes *E. coli* DH5α *library* (Invitrogen) mediante choque térmico y se sembraron en placas de LB-Agar suplementado con 50 µg/ml de kanamicina. Tras una noche de incubación a 37° C se analizaron 10 colonias de cada construcción mediante PCR de colonia utilizando como oligonucleótidos T7 SH de secuencia 5' TAATACGACTCACTATAGG 3' y pET28b-scFvMehtaRv según el apartado 4.3.

Se inoculó una colonia positiva de cada construcción en LB suplementado con kanamicina y se cultivó a 37° C y 250 rpm durante toda la noche. Los vectores contruidos se purificaron con el kit Wizard (Promega) y se verificó la secuencia

mediante secuenciación de DNA. Una vez purificados y verificados los vectores, se transformaron células competentes *E. coli* BL21 (DE3). Se inoculó una colonia en 10 ml de LB suplementado con 50 µg/ml de kanamicina y se creció a 37° C y 250 rpm durante toda la noche. La expresión de las scFv con IPTG se llevó a cabo según el apartado 9.1.6. Se centrifugaron los cultivos y se resuspendieron los precipitados celulares en Bugbuster (Novagen) a 1/20 (v/v) de cultivo suplementado con 125 U de benzonasa (Novagen), 5 KU de lisozima (Novagen) e inhibidores de proteasas 1x (Complete, Roche). Se incubaron en rotación durante 20 minutos a temperatura ambiente y se centrifugaron a 13.500 rpm durante 10 minutos a 4° C, separando así la fracción soluble de la insoluble. Los niveles de expresión se analizaron en geles de poliacrilamida-SDS al 10% mediante tinción con Azul brillante de *Coomassie* (Bio-Rad).

Las scFv se purificaron mediante cromatografía de afinidad en columnas de Ni²⁺ y el ÄKTA Prime (ver apartado 8.3). Para una columna Ni²⁺ de 5ml se inyectaron 20 ml de fracción soluble del cultivo celular. Todas las muestras fueron filtradas (Sartorius, 0,45 µm) antes de la inyección. La pureza de las scFv purificadas en las fracciones recogidas se comprobó mediante electroforesis en geles de poliacrilamida-SDS 10% y tinción con Azul brillante de *Coomassie*. Las fracciones se juntaron, se dializaron frente a PBS durante toda la noche a 4° C para eliminar el imidazol y se conservaron a -80° C.

9.2.4 Expresión y purificación de las scFv en pSANG10-3F

El vector pSANG10-3F fue cedido por el Dr. Schofield (*Wellcome Trust Sanger Institute*, Cambridge, UK) (Martin *et al*, 2006). Este vector deriva del vector pET26b(+) al que se introdujo la secuencia codificadora del tag 3xFLAG entre los sitios de corte *Bgl*I (237) y *Xho*I (158) y el tag 6xHis en el extremo carboxilo-terminal del sitio de clonaje de las scFv (Figura 10).

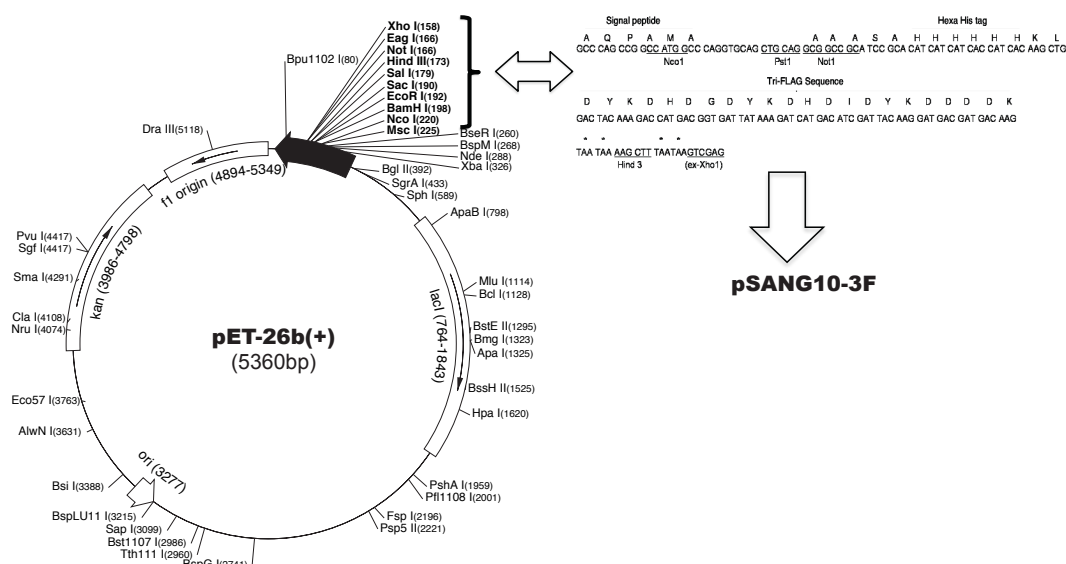


Figura 10. Construcción del vector pSANG10-3F

Los vectores derivados de pSANG10-3F y pET28b conteniendo las secuencias codificantes de las scFv se digirieron con *NotI* (Fermentas) y *NcoI* (Fermentas) durante 2h a 37° C y se ligaron con T4 ligasa según el apartado 9.1.9. Se transformaron células competentes *E. coli* DH5α *library* mediante choque térmico y se sembraron en placas de LB Agar suplementado con 50 µg/ml de kanamicina. Tras una noche de incubación a 37° C se analizaron 10 colonias de cada construcción mediante PCR de colonia utilizando como oligonucleótidos directo y reverso T7 SH y T7 Ter, de secuencia 5 'GCTAGTTATTGCTCAGCGG3', obteniendo un porcentaje de colonias positivas del 100%. Se inoculó una colonia positiva de cada construcción y se cultivó a 37° C y 250 rpm durante toda la noche. Los vectores construidos se purificaron mediante miniprep con el kit Wizard, se verificó la secuencia mediante secuenciación de DNA y se transformaron células competentes BL21 (DE3) mediante choque térmico. La expresión y purificación de las scFv se llevó a cabo siguiendo el protocolo descrito en el apartado anterior.

9.2.5 ELISA de las scFv fusionadas al 3xFLAG

Se analizó mediante ELISA la capacidad de reconocimiento del antígeno por parte de las scFv fusionadas al tag 3xFLAG. Para ello, placas de 96 (Falcon) se recubrieron con 0,3 µg de proteína recombinante purificada o de tag por pocillo durante toda la noche a 4° C. Los pocillos se lavaron con PBS y se bloquearon con PBS-Molico al 2%. Se añadió el periplasma diluido 1/2 en PBS-Tween 20 al 0,1% y Molico al 2%.

Como anticuerpo secundario se utilizó el anticuerpo monoclonal anti FLAG (Sigma) a una dilución 1:1.000 y posteriormente anti IgG de ratón conjugado a peroxidasa (Sigma) (dilución 1:10.000). Como sustrato revelador se utilizó TMB y H₂SO₄ 1M para detener la reacción.

9.2.6 Expresión y purificación de las scFv fusionadas al 3xFLAG en pET32b

El vector pSANG10-3F/scFv-Mehta conteniendo las regiones codificantes de las scFv y el vector pET32b (Novagen) se digirieron con *NcoI* y *HindIII* (Fermentas) durante 2h a 37° C. La ligación de los insertos en el vector, la selección de las colonias positivas, la expresión y purificación de las scFv en pET32b se realizó tal y como se describe en el apartado 9.2.3.

10. Análisis de las constantes cinéticas de las scFv seleccionadas

El estudio de las constantes de afinidad de las scFv se realizó con un equipo BIAcore X (BIAcore). Se utilizaron chips CM5 para el análisis de la interacción entre las scFv y el antígeno frente al que fueron seleccionadas Tanto las proteínas recombinantes como los controles (MBP y GST) se unieron covalentemente a la superficie del chip. La activación de los grupos carboxilo de la matriz de dextrano del chip se llevó a cabo a un flujo de 10 ml/min inyectando EDC (1-etil-3-(3-dimetilaminopropil)-carbodiimida) 0,4 M y NHS (N-hidroxisuccinimida) 0,1 M en una relación 1:1 (v/v) durante 6 min, generando ésteres succinimida activos que reaccionan espontáneamente con los grupos amino primarios de los ligandos, fijándolos covalentemente a la matriz (Figura 12). Para bloquear los grupos ésteres activos y evitar que fijen a la matriz los analitos (moléculas que se inyectan en las Fc para estudiar la cinética de su unión con el ligando) así como para eliminar de la superficie los ligandos que hayan quedado electrostáticamente unidos, se inyectó cloruro de etanolamina 1M en fosfato sódico 0,1 M pH 7.0 (Biacore) en el chip durante 6 min al mismo flujo.

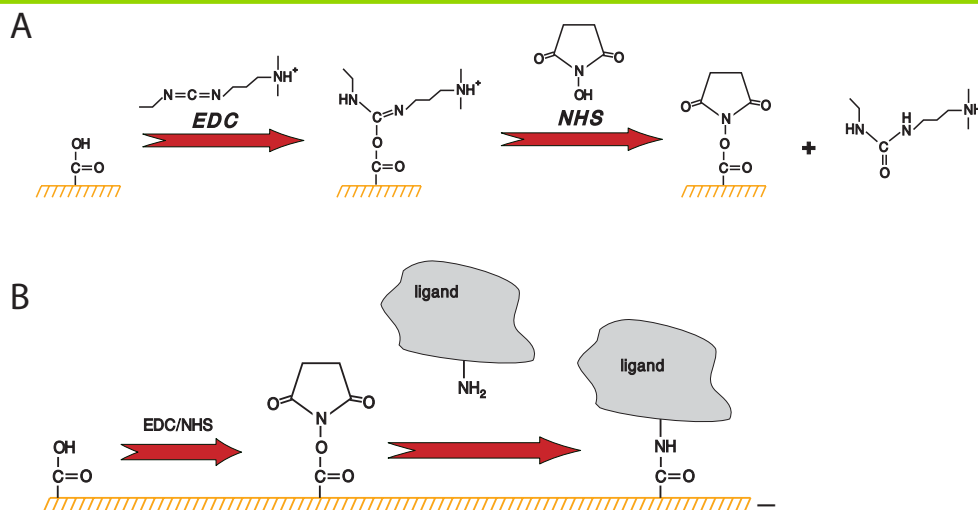


Figura 12. Fijación covalente de los ligandos al chip a través de los grupos amino. Tras la activación del chip con los compuestos EDC y NHS se generan ésteres succinimida activos (A) capaces de unir covalentemente a la matriz los ligandos a través de sus grupos amino (B).

Las proteínas se fijaron en el chip a una concentración de 50 $\mu\text{g/ml}$ en tampón acetato 10 mM pH 4,0. El pH del tampón de fijación debe estar entre 0,5 y 1 unidades por debajo del punto isoeléctrico de la proteína que se quiere unir al chip y alrededor de 3,5, pH en el que la superficie de dextrano está cargada negativamente. Con un tampón de pH entre 3,5 y el punto isoeléctrico del ligando se consigue una preconcentración del ligando sobre la superficie del chip debida a la atracción electrostática entre el ligando, cargado positivamente, y la superficie, cargada negativamente. Con un pH 4 todas las proteínas a fijar en el chip se encuentran por debajo de su punto isoeléctrico.

En la Fc1 de cada chip se fijó MBP o GST, dependiendo del *tag* de cada antígeno, como control negativo de unión de las scFv; en la Fc2 se fijó proteína recombinante utilizada como antígeno de selección de anticuerpos. El flujo utilizado en la fijación fue de 10 $\mu\text{l/min}$ y el tampón PBS suplementado con Surfactante p20 al 0.005% (BIAcore). La cantidad de ligando que se fija dependerá principalmente de la masa molecular del ligando y de los analitos que se quieran estudiar ya que la respuesta SPR es directamente proporcional a la masa que hay en la superficie del chip. Para el análisis cinético se recomienda fijar la menor cantidad de ligando que dé lugar a una respuesta medible del analito. La capacidad de unión del analito al chip se relaciona con la cantidad de ligando fijada a la superficie mediante la fórmula:

$$\text{Capacidad de unión del analito (RU)} = \frac{\text{Peso molecular del analito (kDa)}}{\text{Peso molecular del ligando (kDa)}} \times \text{Ligando inmovilizado (RU)}$$

Teniendo en cuenta que la fijación del ligando genera diferentes orientaciones de la proteína sobre el chip, se estima que tan sólo 1/3 del ligando fijado expone el epítipo correctamente a los analitos. Por lo tanto, fijando una cantidad de ligando que genere una respuesta máxima teórica del analito de 1000 RU tendremos una respuesta real máxima (suponiendo una interacción 1:1) de 333 RU. Fijamos en la Fc2 la cantidad de antígeno recombinante necesaria para tener una respuesta máxima teórica de las scFv de 1000 RU.

Para el análisis de las constantes de afinidad de las scFv se utilizó un flujo de 20 µl/min y se inyectaron diferentes diluciones de los anticuerpos durante 2 minutos. Las diluciones utilizadas fueron de 1, 0.5, 0.25, 0.125 y 0.00625 µM. Una vez obtenidos los sensorgramas, las constantes de afinidad se calcularon utilizando el programa BIAsoftware. El modelo elegido fue el de Langmuir 1:1 dado que las scFv presentan un único lugar de reconocimiento del antígeno.

11. Obtención de anticuerpos monoespecíficos frente a fragmentos proteicos (PrEST)

11.1 Diseño de los fragmentos proteicos PrEST

El diseño de los PrEST se llevó a cabo en tres pasos. Se evaluó el grado de hidrofobicidad de la secuencia completa de las cinco proteínas seleccionadas (AGT, EPHB3, PLAU, PTK7 y SPARC) con el algoritmo desarrollado por Kyte y Doolittle (1982) para evitar regiones hidrofóbicas que favorecerían la insolubilidad de la proteína al ser expresada en bacteria. La predicción de las regiones transmembrana se llevó a cabo mediante el algoritmo SOSUI (Hirokawa *et al*, 1998). Finalmente, se hizo un BLAST de los fragmentos seleccionados frente a la base de datos del NCBI (www.ncbi.nlm.nih.gov/blast/Blast.cgi) para escoger los fragmentos que mostrasen menor homología frente al resto de proteínas humanas, evitando así posibles reacciones cruzadas a la hora de utilizar los anticuerpos obtenidos frente a dichos fragmentos.

11.2 Expresión y purificación de fragmentos PrEST

Los fragmentos seleccionados se amplificaron mediante PCR a partir de clones de secuencia completa de la MGC. Las amplificaciones se llevaron a cabo con la enzima Platinum Pfx DNA polimerasa (Invitrogen). En los extremos de los oligonucleótidos utilizados (Tabla 4) se añadieron dianas de restricción para las enzimas

NdeI en el extremo 5' y *BamHI* en el extremo 3' para llevar a cabo el clonaje en pET15b. Las reacciones de amplificación se realizaron a partir de 20 ng de vector con una concentración final de cada nucleótido de 0,4 μ M, dNTP 0,2 μ M, $MgCl_2$ 1,5 μ M, tampón 1X y 2,5 U de Pfx DNA polimerasa en un volumen final de 50 μ l. Como programa de amplificación se utilizó 1 ciclo de 3 min a 94° C, 25 ciclos (15 s a 94° C, 30 s a 55° C y 1 min/Kb a 69° C) y un ciclo de 7 min a 68° C.

Tabla 4. Oligonucleótidos utilizados para amplificar los feragmentos PrEST

Gen	Oligonucleótido	Secuencia
AGT	NdeIAGT-345	CTGCCTGCATATGATCCAGCCTCACTATGCCTCTG
	AGT-461BamHI	GATCATAGGATCCCTACAGGAATGGGCGGTTTCAGG
EPHB3	NdeIEPHB3-223	CTTCGCACATATGCCGAGACCCTCACTGGGGCGG
	EPHB3-340BamHI	CCTCGGGATCCCTATGGCACGGTGGTACAGGCAC
PLAU	NdeIPLAU-25	GCAGCAATCATATGCATCAAGTTCCATCGAACTG
	PLAU-171BamHI	CGGGGGATCCCTACTTTTGGCCACACTGAAATTT
PTK7	NdeIPTK7-728	GTTCTACCATATGAAGCGCTGCAAAGCCAAGCGGC
	PTK7-830BamHI	CATCCTTGCTGGATCCCTACTTCACAAGTACCAGG
SPARC	NdeISPARC-178	GCTCAAGCATATGCTGGTCAACCCTGTATGAGAGG
	SPARC-301BamHI	GCTTAGGATCCCTAATCCTTGTCGATATCCTTCTG

Los productos de PCR y el vector pET15b (Novagen) se digirieron con *NdeI* y *BamHI* (Fermentas) durante 2h a 37° C en los tampones recomendados por el fabricante. Se eligió el vector pET15b porque sólo contiene el *tag* de 6xHis. Los insertos y vector digeridos se ligaron con T4 Ligasa utilizando 20 ng de vector por ligación y una proporción molar vector: inserto 1:3. Se introdujo mediante transformación por choque térmico de 2 μ l del producto de ligación en 20 μ l de células *E. coli* TOP10 (Invitrogen) y se plaquearon en placas de LB-Agar suplementadas con 100 μ g/ml de ampicilina.

La presencia de los insertos en el vector se analizó mediante PCR de 10 colonias utilizando como oligonucleótido directo el oligonucleótido T7 SH y como oligonucleótido reverso el oligo específico de cada fragmento, el mismo utilizado para obtener la secuencia seleccionada a partir del clon de la MGC (Tabla 3). Se secuenciaron los vectores obtenidos para verificar la presencia de los insertos.

Se transformaron células competentes de *E. coli* BL21 (DE3) con los vectores pET15b-PrEST. La expresión de los fragmentos PrEST se llevó a cabo según los protocolos descritos anteriormente (apartado 6).

Una vez inducida la expresión de las proteínas en 1L de cultivo, estos se centrifugaron a 4.000 rpm durante 15 min. Los precipitados se resuspendieron en 0,05 volúmenes de tampón fosfato sódico 20 mM pH 7,4, NaCl 0,5 M, urea 6 M e imidazol 20 mM y se dejaron durante toda la noche a 4° C en rotación para lisar las células. Se centrifugaron durante 10 minutos a 4000 rpm y los sobrenadantes se sonicaron 3 veces durante 30 segundos. Los niveles de expresión proteica se analizaron mediante electroforesis de los extractos proteicos totales en gel de poliacrilamida-SDS al 12.5% y tinción con Azul brillante de *Coomassie* (Bio-Rad).

Los PrEST fusionados al *tag* 6xHis se purificaron mediante cromatografía de afinidad IMAC utilizando columnas de Ni²⁺ HiTrap™ Crude (Amersham Biosciences) de 5 ml. Se inyectaron 20 ml de extracto total desnaturalizados en urea (tampón fosfato sódico 20 mM pH 7,4, NaCl 0,5 M, urea 6 M e imidazol 20 mM) en la columna y se purificaron según el apartado 8.3. El tampón de equilibrado fue el mismo utilizado para la lisis de los precipitados celulares y como tampón de elución se utilizó el tampón de equilibrado con una concentración final de imidazol de 0.5 M. Las fracciones de elución que contenían los PrEST se dializaron frente a PBS suplementado con urea 1 M durante toda la noche a 4° C y se conservaron a 4° C.

11.3 Obtención de sueros policlonales

La inmunización de los conejos se llevó a cabo en la empresa Vivotecnia. Se inmunizaron conejos hembra *New Zealand White* utilizando como antígenos los PrEST diluídos en PBS hasta una concentración de urea 1M. La primera inmunización se realizó inoculando por vía intradérmica 250 µg de antígeno en adyuvante completo de Freund. Se dieron 4 dosis de recuerdo con 125 µg de antígeno en adyuvante incompleto de Freund.

11.4 ELISA de sueros policlonales

Se fijaron 0,3 µg de antígeno por pocillo en placas de 96 pocillos (Falcon) durante toda la noche a 4° C en PBS. Los pocillos se bloquearon con PBS Tween 20 0,05% y leche en polvo semidesnatada al 2% durante 2h a temperatura ambiente. Los sueros se añadieron a los pocillos haciendo diluciones seriadas de factor 1/5 a partir una dilución inicial 1/50 en PBS Tween 20 0,05% y leche en polvo semidesnatada al 2% y se incubaron durante 1h a 37° C. Los pocillos se lavaron tres veces con PBS Tween 20

0,05% y se añadió Proteína A conjugada a peroxidasa (Sigma) a una dilución 1:5000 como anticuerpo secundario y se incubó durante 1h a temperatura ambiente. Se lavaron los pocillos tres veces con PBS-Tween 20 0,05%, se añadieron 100 µl de TMB y se detuvo la reacción con 100 µl de H₂SO₄ 1M. La absorbancia de cada pocillo se midió a una longitud de onda de 450 nm en un lector Model 550 (Bio-Rad).

11.5 Purificación de anticuerpos monoespecíficos

Se eligió el suero de mejor título para purificar los anticuerpos monoespecíficos. Se utilizaron columnas HiTrap NHS-activated HP 1ml (Amersham Biosciences) para construir una columna de afinidad específica para cada PrEST. Para la fijación de las proteínas a la resina se diluyeron los PrEST purificados en tampón NaHCO₃ 0.2M pH 8, NaCl 0,5M a una concentración final de 0,5 µg/µl. Como tampones de bloqueo de la columna se utilizó etanolamina 0,5 M pH 8,3, NaCl 0,5 M y acetato sódico 0,1 M pH 4,0, NaCl 0,5 M. Una vez fijadas las proteínas se neutralizó el pH de las columnas con el tampón de equilibrado fosfato sódico 20 mM pH 7,0.

Los anticuerpos monoespecíficos se purificaron en un equipo ÄKTA prime. Se inyectaron 5 ml de suero en la columna. Como tampón de elución se utilizó glicina-HCl 0,1 M pH 2,7. Brevemente, el programa utilizado para la purificación de los anticuerpos consta de un paso de equilibrado de la columna con 10 ml de tampón de equilibrado a un flujo de 1 ml/min, inyección de la muestra y elución de los anticuerpos con 10 ml de tampón de elución a un flujo de 1 ml/min. Los anticuerpos eluyen en las fracciones 2 y 3 principalmente.

Las fracciones se recogieron en 100 µl de Tris-HCl 1 M pH 9,5 para neutralizar el pH. Los anticuerpos se dializaron frente a PBS durante toda la noche a 4° C. Se analizó su capacidad de reconocimiento del antígeno mediante ELISA.

12. Producción de anticuerpos monoclonales

La producción de anticuerpos monoclonales específicos frente a BMP4 y LCN2 se llevó a cabo en la Unidad de Anticuerpos Monoclonales del Programa de Biotecnología del CNIO. Se utilizaron como antígenos la proteína BMP4 de secuencia parcial fusionada a MBP y LCN2 de secuencia completa fusionada a GST, respectivamente y ratones Balb/C. Se llevaron a cabo 5 inmunizaciones

intreperitoneales con 50 µg de antígeno acompañado de adyuvante completo de Freund en la primera y tercera inyección. Los linfocitos B del bazo de los ratones se fusionaron a células de mieloma NS-1 (P3/NS1/1-Ag4-1) utilizando la técnica convencional (Mason *et al*, 1983).

El sobrenadante de los hibridomas se analizó inicialmente mediante ELISA frente a las proteínas purificadas, inmunocitoquímica y *western blot*. La inmunocitoquímica se llevó a cabo mediante transfección con fosfato cálcico de células Cos-1 (riñón de mono verde africano, DSMZ) con los vectores BMP4-pcDNA3.1/nV5-DEST y LCN2-pcDNA3.1/nV5-DEST (construidos con el sistema Gateway, apartado 5). La transfección con fosfato cálcico genera precipitados de DNA que son internalizados por las células. Se mezclaron 30 µg de vector purificado con CaCl₂ 2,5 M en un volumen final de 500 µl. esta mezcla se añadió a 500 µl de HEPES 2x agitando levemente y se dejó en reposo durante 20 min a temperatura ambiente para que se formasen los precipitados. La mezcla se añadió a las células en fase de crecimiento exponencial y se dejó durante 6 h en un incubador humidificado a 37° C y un 5% de CO₂. Finalmente se retiró la mezcla de las placas de cultivo, se lavaron las células con PBS y se añadió medio fresco DMEM (Sigma) suplementado con suero fetal bovino al 10% y penicilina y estreptomicina al 1%.

13. Western blot con muestras de tejido

Tejidos congelados procedentes de 5 pacientes en estadio de Dukes de A a C se lavaron dos veces con PBS y se extrajeron las proteínas tanto de tejido sano normal como de tejido tumoral mediante tres pulsos de sonicación (Digital Sonifier, Branson) de las muestras en tampón de lisis (Tris-HCl 50 mM pH 7,4, NaCl 150 mM, DTT 2 mM, SDS 0,1%). Los extractos proteicos se resolvieron por electroforesis en geles de poliacrilamida-SDS al 12% y se transfirieron a membranas de nitrocelulosa (Hybond-C Extra, Amersham Biosciences). Tras bloquear las membranas con PBS con leche en polvo semidesnatada al 5% se añadieron los anticuerpos primarios. Los sobrenadantes de los hibridomas frente a BMP4 y LCN2 se utilizaron directamente, sin diluir; los anticuerpos policlonales frente a ANXA3 se utilizaron en una dilución 1:1000; antiSPARC (Novocastra) se utilizó a dilución 1:1000; antiSPP1 (Abcam) a 1:10000;

antiMMP7 (Chemicon) a 1:2000 y antiMMP11 (Chemicon) a 1:200. Como control de carga se utilizó el anticuerpo anti Tubulina (Sigma) a una dilución 1:5000. Tras la incubación con los anticuerpos primarios, las membranas se lavaron 3 veces con PBS-Tween 20 al 0,05% y se incubaron durante 90 minutos con los anticuerpos secundarios correspondientes: anti IgG de ratón conjugado con peroxidasa (DakoCytomation) a una dilución 1:1000 o anti IgG de conejo conjugado a Peroxidasa (Sigma) a una dilución 1:1000. La unión de los anticuerpos se detectó con el reactivo ECL (Amersham biosciences) o SuperSignal Femto (Pierce).

14. Preparación y análisis de *microarrays* de tejido

Para preparar los *microarrays* de tejido de CCR se utilizaron muestras tumorales de CCR pertenecientes a pacientes en estadio de Dukes A, B o C, muestras de tejido normal de los mismos pacientes y muestras de tejido control (lengua, mama, hígado, pulmón, páncreas y placenta). Las muestras fueron recolectadas y clasificadas de forma anónima por el Banco de tumores (CNIO). Los bloques de parafina utilizados se seleccionaron en base a su disponibilidad, debiendo de contener muestra suficiente para un mínimo de 60 secciones de 1mm de diámetro. Se utilizó un *arrayer* de tejido (Beecher Instruments) para construir dos bloques de *microarrays* de tejido con un total de 97 muestras, incluyendo réplicas de algunas muestras y diferentes localizaciones de la neoplasia a lo largo de la región colorrectal. El grado de expresión de las proteínas en el *microarray* se valoró de 1 a 3 (1 si la tinción es menor del 10%, 2 si la tinción es del 10 al 50% y 3 si la tinción es mayor del 50%).

14.1 Tinción con scFv

En el caso de la scFv antiSPARC, el *microarray* se incubó con la scFv purificada a una dilución 1:25. Para detectar la unión de la scFv se utilizó el anticuerpo anti FLAG (Sigma) a una dilución 1:200. La tinción específica se detectó mediante la incubación con anti IgG de ratón conjugado a Biotina. La visualización de la interacción específica se llevó a cabo mediante el sistema EnVision HRP (Dako) siguiendo las instrucciones del fabricante. Brevemente, se bloqueó la actividad peroxidasa endógena con agua oxigenada al 3% durante 5 min a temperatura ambiente y se lavó la preparación con PBS. Las proteínas endógenas se bloquearon con albúmina durante 10 min a temperatura ambiente. El anticuerpo primario se incubó durante 30 min a temperatura

ambiente y, tras lavar la muestra con PBS, se añadió el polímero marcado (anticuerpo secundario + HRP) y se incubó durante 30 min a temperatura ambiente. A continuación se añadió el substrato cromógeno DAB (3, 3'-diaminobenzidina tetrahidroclorido) y se incubó durante 5 min a temperatura ambiente. Finalmente, tras lavar con agua destilada, se contratiñó la preparación con hematoxilina. Las muestras se deshidrataron con etanol y xilol antes de montar las preparaciones.

14.2 Tinción con anticuerpos monoespecíficos

Los anticuerpos monoespecíficos anti PTK7 y anti SPARC se utilizaron a una dilución 1:50. A continuación se realizó la incubación con anti IgG de conejo conjugado a Biotina. La visualización de la interacción específica se llevó cabo mediante el sistema LSAB-HRP (Dako). Brevemente, se bloquearon la actividad peroxidasa endógena y las proteínas de la muestra añadiendo agua oxigenada al 3% sobre la muestra durante 5 min a temperatura ambiente. La preparación se lavó con PBS, se añadió el anticuerpo primario y se incubó durante 30 min a temperatura ambiente. Tras lavar con PBS se añadió el anticuerpo secundario, se incubó durante 30 min a temperatura ambiente y se lavó nuevamente con PBS. Se añadió la enzima peroxidasa y se incubó durante 25 min a temperatura ambiente. A continuación se añadió la solución de substrato cromógeno DAB y se incubó durante 10 min a temperatura ambiente.

14.3 Tinción con anticuerpos mono y policlonales

En el caso de los anticuerpos mono y policlonales, los *arrays* fueron incubados con las siguientes diluciones: antiANXA3 (1:500), -BMP4 (1:25), -LCN2 (1:2), -SPARC (1:25), -SPP1 (1:750), -MMP7 (1:200) y antiMMP11 (1:50). Como control de proliferación se incluyó la tinción con antiKi67 (DakoCytomation) a una dilución 1:50. La unión específica de cada anticuerpo primario se detectó mediante incubación con anti IgG de ratón o anti IgG de conejo conjugados a Biotina. La visualización de las interacciones específicas se llevó cabo mediante el sistema EnVision-HRP (Dako).

15. Trasfección de siRNA

Se utilizaron siRNA (Quiagen) dirigidos frente a los mRNA de los genes ANXA3, BMP4 y MMP11 (Tabla 5) para inhibir la expresión de dichos genes y analizar su efecto en los niveles de apoptosis de las células.

Tabla 5. Secuencias diana de los siRNA utilizados

Gen	Secuencia diana
ANXA3	CAGGACAAGCAGGCAAATGAA TTCCCTATATTCAGCAATTAA
BMP4	CACGAGACTGGTCCACCACAA GCGAGCCATGCTAGTTTGATA
MMP11	TCCCGAGAAGAACAAGATCTA Validado

Los siRNA se transfectaron en células Hct15 a un 70% de confluencia. Antes de la transfección se retiró el medio de cultivo normal de las células y se añadió Optimem (Gibco) sin antibióticos ni suero fetal bovino. Se añadieron 5 µl de Oligofectamina (Invitrogen) y los siRNA a una concentración final de 0,2 µM en 500 µl de Optimem. Tras 1h de incubación a temperatura ambiente se añadió la mezcla a las células y se incubaron 24h a 37° C. Pasado este tiempo se retiró el Optimem y se añadió RPMI suplementado con suero fetal bovino al 10% y antibióticos al 1% (Penicilina/ Estreptomicina) (Gibco). Las células se recogieron a las 24, 48 y 72 h para determinar por PCR cuantitativa (apartado 3.2) el punto de mayor interferencia al que analizar los niveles de apoptosis.

16. Determinación de la apoptosis celular

Tras levantar las células con tripsina-EDTA (Gibco) durante 5 min a 37° C, se resuspendieron en tampón Hepes 10 mM pH 7,4, NaCl 140 mM, CaCl₂ 2,5 mM a una concentración de 3x10⁶ células/ml. Se prepararon por duplicado tubos Facs (Falcon) con 100 µl de células y se añadieron 5 µl de anticuerpo frente a ANXAV (BD) y 10 µl de ioduro de propidio a 1mg/ml (Sigma). Tras incubar la mezcla durante 15 min a temperatura ambiente en oscuridad se añadieron 400 µl del tampón anterior y se analizaron los niveles de apoptosis por citometría de flujo en un equipo BD FACSCanto II (BD Biosciences). Se contaron 20.000 eventos con el software FACSDiva v 5.0 (BD Biosciences).

17. Ensayo de proliferación celular

17.1 Construcción de los vectores pEGFP/C1-ANXA3/BMP4

Se amplificaron los cDNA de los genes ANXA3 y BMP4 a partir de los clones de la colección MGC por PCR con Platinum Pfx DNA polimerasa (Invitrogen) en un volumen final de 50 μ l (apartado 11.2). Los oligonucleótidos utilizados para la amplificación por PCR se muestran en la Tabla 6. Se digirieron los productos de PCR y el vector pEGFP/C1 (BD Biosciences Clontech) con las enzimas de restricción *XhoI* y *HindIII* durante 2h a 37° C. Los productos de PCR y el vector digeridos se ligaron y transformaron en las condiciones habituales y se plaquearon en placas de LB-Agar suplementadas con 50 μ g/ml de kanamicina. Las colonias positivas se chequearon por PCR de colonia utilizando como oligonucleótidos los específicos de cada gen. A partir de la placa réplica se inoculó una colonia positiva de cada construcción en 10 ml de LB suplementado con 50 μ g/ml de kanamicina y se purificaron los vectores por miniprep utilizando el kit Wizard.

Tabla 6. Oligonucleótidos utilizados para subclonar ANXA3 y BMP4 en pEGFP/C1

Gen	Oligonucleótido	Secuencia
ANXA3	ANXA3XhoIFw	AAACTCGAGCATCTATCTGGGTGGACAC
	ANXA3HindIIIRv	AAAAAGCTTTCAGTCATCTCCACCACAGA
BMP4	BMP4XhoIFw	AAACTCGAGCTATTCCTGGTAACCGAATGCT
	BMP4HindIIIRv	AAAAAGCTTTCAGCGGCACCCACATCCCTC

17.2 Ensayo de proliferación celular

Células Hct116 fueron electroporadas con los vectores descritos en el apartado anterior. Para ello, 10^7 células tripsinizadas fueron lavadas 3 veces con PBS para eliminar los restos del medio de cultivo. Tras el último lavado se resuspendieron en PBS y se añadieron 30 μ g de vector (ANXA3-pEGFP/C1, BMP4-pEGFP/C1 y pEGFP/C1) en un volumen final de 800 μ l. A continuación las células fueron electroporadas con un pulso eléctrico de 250 mV, 950 μ F y máxima capacitancia con un electroporador Gene pulse II (Bio-Rad). Tras el pulso eléctrico se añadieron 500 μ l de RPMI suplementado con SFB al 20% y se descartaron las células muertas; las células restantes se sembraron en 5 placas de 96 pocillos (Becton Dickinson) para cada vector a diferentes concentraciones (1000/2000/4000/8000/16000 células) en un volumen final de 100 μ l y

se analizó el número de células vivas cada 24h durante 5 días con MTT ((3-(4,5-dimetiltiazol-2-il)-2,5-difeniltetrazolio bromuro) (Sigma), compuesto que se une a un complejo mitocondrial sólo si la célula está viva. Tras la incubación con MTT durante 3h, se retiró el medio, se añadieron 100 µl de DMSO (Dimetil sulfóxido) (Sigma) y se midió la absorbancia en un espectrofotómetro Model 550 (Bio-Rad) a 595 nm.

Resultados

1. Identificación de dianas moleculares de cáncer colorrectal

1.1 Identificación de genes sobreexpresados en cáncer colorrectal mediante el análisis de *microarrays* de DNA

En colaboración con la Unidad de Genómica (CNIO), se analizaron 22 pares de muestras (normal- tumoral) de pacientes con cáncer colorrectal en diferentes estadios para determinar las diferencias en la expresión génica entre tejidos sanos y tumorales; para ello, se utilizaron *microarrays* de cDNA (CNIO Oncochip).

Fueron identificados un total de 1182 genes sobreexpresados. Un total de 371 sondas del chip mostraron una tasa de expresión 1,5 veces mayor en muestras tumorales en comparación con las normales. De estas 371 sondas, 337 corresponden a genes conocidos y 13 a proteínas hipotéticas (Anexo I).

1.2 Selección de proteínas diana a partir de genes alterados en cáncer colorrectal

Se seleccionaron los 50 genes con mayor grado de sobreexpresión (Figura 12).

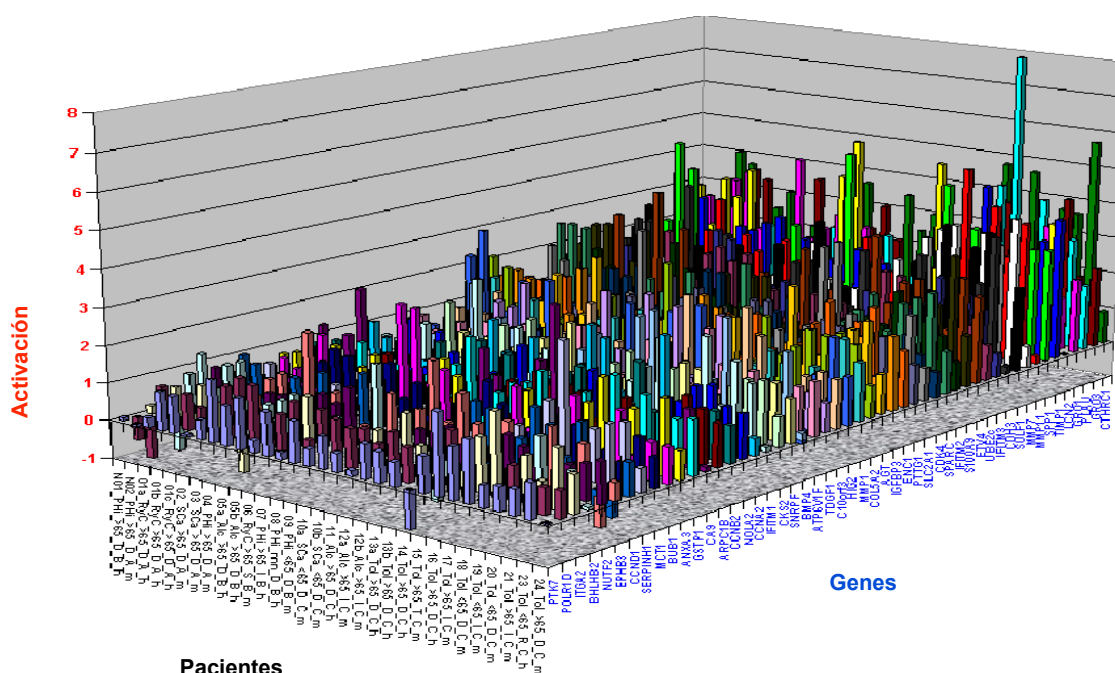


Figura 12. Genes con mayor sobreexpresión en tejidos cáncer colorrectal.

Los genes se clasificaron según la localización celular de su producto proteico utilizando el programa GO (*Gene Ontology Consortium*, www.geneontology.org), donde se describen los genes y sus productos proteicos según su función biológica y

localización celular o tisular; la ontología de cada uno de los 50 genes se muestra en la Tabla 7. Tras dicho análisis, se seleccionaron aquellos genes cuyos productos proteicos se localizasen tanto en la membrana celular como en la matriz extracelular por su mayor accesibilidad a los agentes terapéuticos o diagnósticos. Aplicando estos criterios fueron 29 los genes finalmente seleccionados (Tabla 8). La selección de genes candidatos no estuvo condicionada por características como tamaño o función biológica que desempeñan, de hecho, el rango de los pesos moleculares de las proteínas que codifican oscila entre los 129 kDa de la proteína THBS2 y los 6,9 kDa de HIG2.

Tabla 7. Localización celular de las proteínas sobreexpresadas en cáncer colorrectal

Citosol	Núcleo	Unida a membrana	Transmembrana	Integral de membrana	Extracelular	Circulación	Lumen RE	Mitocondria
ARPC1	BHLHB2	ANXA3	EPHB3	CA9	CTHRC1	AGT	SERPINH1	MRPL50
CKS2	CCND1	LCN2	CDH3	HIG2	COL5A2	IGFB3		
CXCL1	BUB1		CA9	ITGA2	CXCL1	PLAU		
CXCL3	CA9		IFITM1	IFITM1	CXCL3			
BUB1	C10orf3		IFITM2	IFITM2	MMP1			
CDK4	CCNA2		IFITM3	IFITM3	MMP7			
ETV4	CCNB2		PTK7	SLC26A3	MMP11			
ENC1	PTTG1		TDGF1		SPARC			
CCNA2	POLR1D				SPP1			
CCNB2	S100A9				SULF1			
MCT-1	SNRPF				TIMP1			
NUTF2	NOLA2				PLAU			
PTTG1					BMP4			
S100A9					TDGF1			
UBE2C					THBS2			
					SEPP1			

En rojo se marcan las regiones celulares en las que se focalizó el estudio

1.3 Validación por PCR cuantitativa de las dianas seleccionadas

Para confirmar los valores de expresión obtenidos con los *microarrays* de cDNA, se analizaron los niveles de expresión del RNA mensajero correspondientes a 15 muestras tumorales (Figura 13) y en 8 líneas de cáncer de colon (CaCo2, Co115, Colo205, Hct15, Hct116, RKO, Sw48 y Sw480) (Figura 14) mediante PCR cuantitativa para confirmar los valores de expresión obtenidos con los *microarrays* de cDNA. Como referencia se utilizó un *pool* de cDNA generado a partir de muestras de colon normales.

Se analizaron 9 genes seleccionados del estudio genómico de partida, SPP1, BMP4, CXCL1, CXCL3, MMP11, MMP7, SPARC, ANXA3 y LCN2. Los resultados obtenidos muestran un patrón muy distinto en el perfil de expresión de los marcadores seleccionados entre las muestras tumorales y las líneas celulares. El análisis cuantitativo de los tumores se observa que los genes LCN2, CXCL3, MMP11 y MMP7 presentan un perfil de expresión muy similar en todos los tumores analizados independientemente del estadio Dukes en el que se encuentran mientras que los genes BMP4, SPARC y ANXA3

son muy heterogéneos en cuanto a su patrón de expresión génica en los tumores analizados. Las muestras tumorales muestran mayor grado de sobreexpresión de los genes seleccionados que las líneas celulares. Exceptuando los casos de ANXA3, BMP4 y MMP11, que aparecen sobreexpresados en casi todas las líneas celulares, las líneas analizadas no presentan sobreexpresión del resto de los marcadores seleccionados para el estudio. Esto parece indicar que los estudios genómicos que parten de muestras tumorales humanas reflejan no sólo el grado de expresión de las células tumorales sino también del estroma que las rodea. A la vista de los resultados podemos concluir que la mayor parte de los marcadores seleccionados son propios de las células estromales. Otra explicación de los resultados obtenidos es que como las proteínas diana, aunque sintetizadas por las células tumorales propiamente dichas, forman parte de la matriz extracelular, su expresión puede estar condicionada a la presencia de factores paracrinós ausentes en los medios de cultivo, por lo que su expresión no se ve alterada.

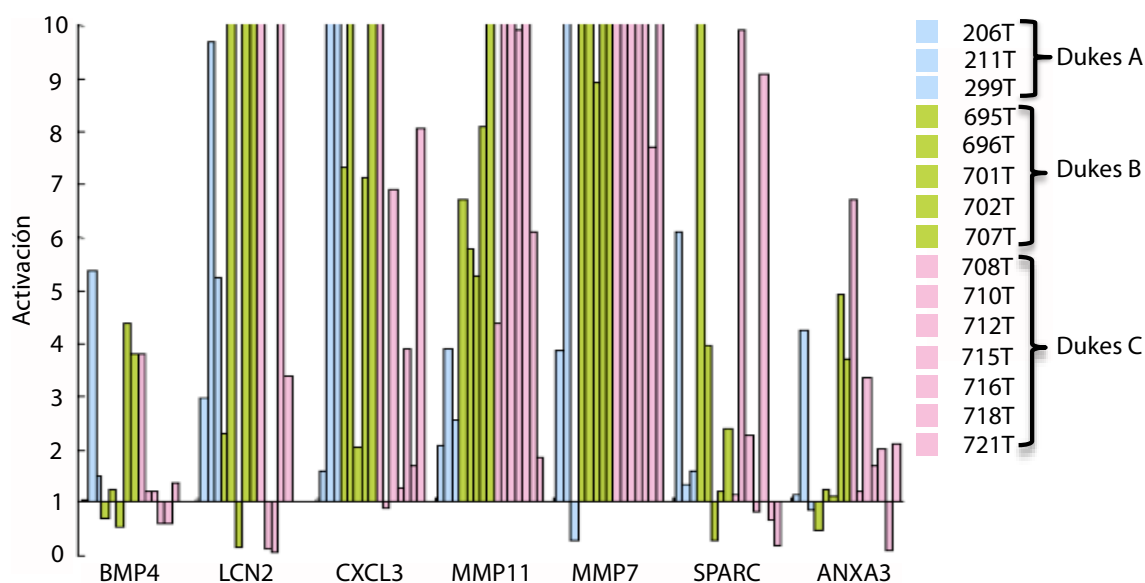


Figura 13. Análisis de la expresión génica mediante Q RT PCR en nuestras tumorales. Se analizan los niveles de expresión de 7 dianas genómicas en 15 muestras tumorales de cáncer colorectal. Las muestras de color azul pertenecen a tumores en estadio Duke's A, las muestras de color verde pertenecen a tumores en estadio B y las muestras de color rosa a tumores en estadio Duke's C

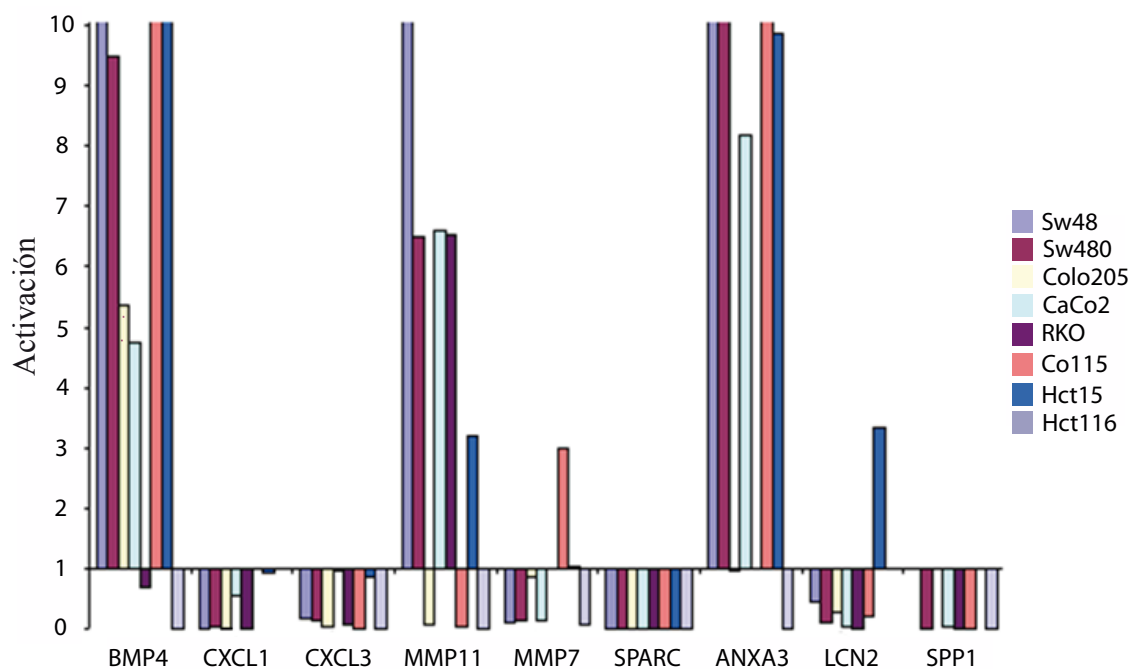


Figura 14. Análisis de la expresión génica mediante Q RT PCR en líneas celulares. Se analizarán los niveles de expresión de 9 dianas genómicas en líneas celulares de cáncer colorrectal. Tan sólo tres de ellas, BMP4, MMP11 y ANXA3 aparecen sobreexpresadas en las líneas celulares analizadas.

2. Producción de proteínas recombinantes a partir de los genes seleccionados

Para la obtención del producto proteico de los genes seleccionados es necesario disponer de los clones de cDNA correspondientes a partir de los cuales construir los vectores de expresión en *Escherichia coli*. Elegimos *E. coli* por ser el sistema de expresión de proteínas más adecuado a la hora de llevar a cabo estudios a gran escala dado su facilidad de manejo, alta velocidad de crecimiento y bajo coste. Para la construcción de los vectores de expresión utilizamos el sistema *Gateway* por la rapidez con la que se obtienen los vectores dado que en su construcción no se utiliza la digestión con enzimas de restricción ni la ligación con ligasas y por su enorme versatilidad ya que a partir de un único vector de entrada se pueden construir todos los vectores destino. Dentro del sistema *Gateway*, escogimos el sistema TOPO por presentar ventajas sobre el sistema DONOR. Los oligonucleótidos utilizados en la amplificación de los cDNAs de interés son de menor tamaño en el caso de TOPO por lo que se introducen menos residuos foráneos en la secuencia amplificada.

2.1 Expresión de proteínas a partir de clones de cDNA de secuencia parcial

Como fuente de los cDNAs para la expresión de las proteínas se utilizaron inicialmente los mismos clones de cDNA empleados para imprimir el CNIO Oncochip, pertenecientes a la colección IMAGE. Ninguno de los clones impresos en el Oncochip tenía la secuencia completa del cDNA, solamente 11 de los 29 clones IMAGE (38%) contenían al menos un 30% del ORF (Fase de lectura abierta) completo, límite que se estableció para la utilización de los clones como cDNA de partida para la expresión de las proteínas. De estos 11 clones sólo se utilizaron 9 para la expresión de proteínas. Los 2 clones restantes (CA9 e IFITM1) se desecharon por mostrar delecciones en la secuencia tras la amplificación por PCR y clonaje en el vector pENTR/D-TOPO.

Se construyeron un total de 27 vectores de destino tras la recombinación homóloga de los 9 vectores de entrada a *Gateway* con tres vectores destino diferentes: pTH1, pDEST15 y pDEST17. Cuando se transformaron directamente células competentes de *E. coli* de la cepa BL21 (DE3) con los vectores destino, la eficiencia de la recombinación fue menor y más variable (70%) en comparación con la transformación de las células de la cepa DH5 α (95%). Sin embargo el ahorro de tiempo justifica la transformación directa de las células BL21 (DE3) cuando se trabaja con un gran número de proteínas simultáneamente.

En la Figura 15 se muestran los resultados de expresión y solubilidad de las proteínas de secuencia parcial obtenidos mediante geles de poliacrilamida SDS teñidos con Azul brillante de *Coomassie* y *western blot*. En la Tabla 8 se resumen estos resultados. Se obtuvieron como productos solubles todas las proteínas fusionadas a MBP, 7 de las 9 proteínas fusionadas a GST (77.8%) (THBS2, BMP4, SPP1, MMP11, LCN2, IFITM3, IFITM2) y 2 de las 4 proteínas fusionadas a 6xHis (50%) (LCN2 e IFITM2). Estos resultados confirman datos previos que apuntan que la MBP actúa como agente solubilizante a la hora de obtener proteínas solubles en *E. coli*, por encima de GST o 6xHis (Braun *et al*, 2001).

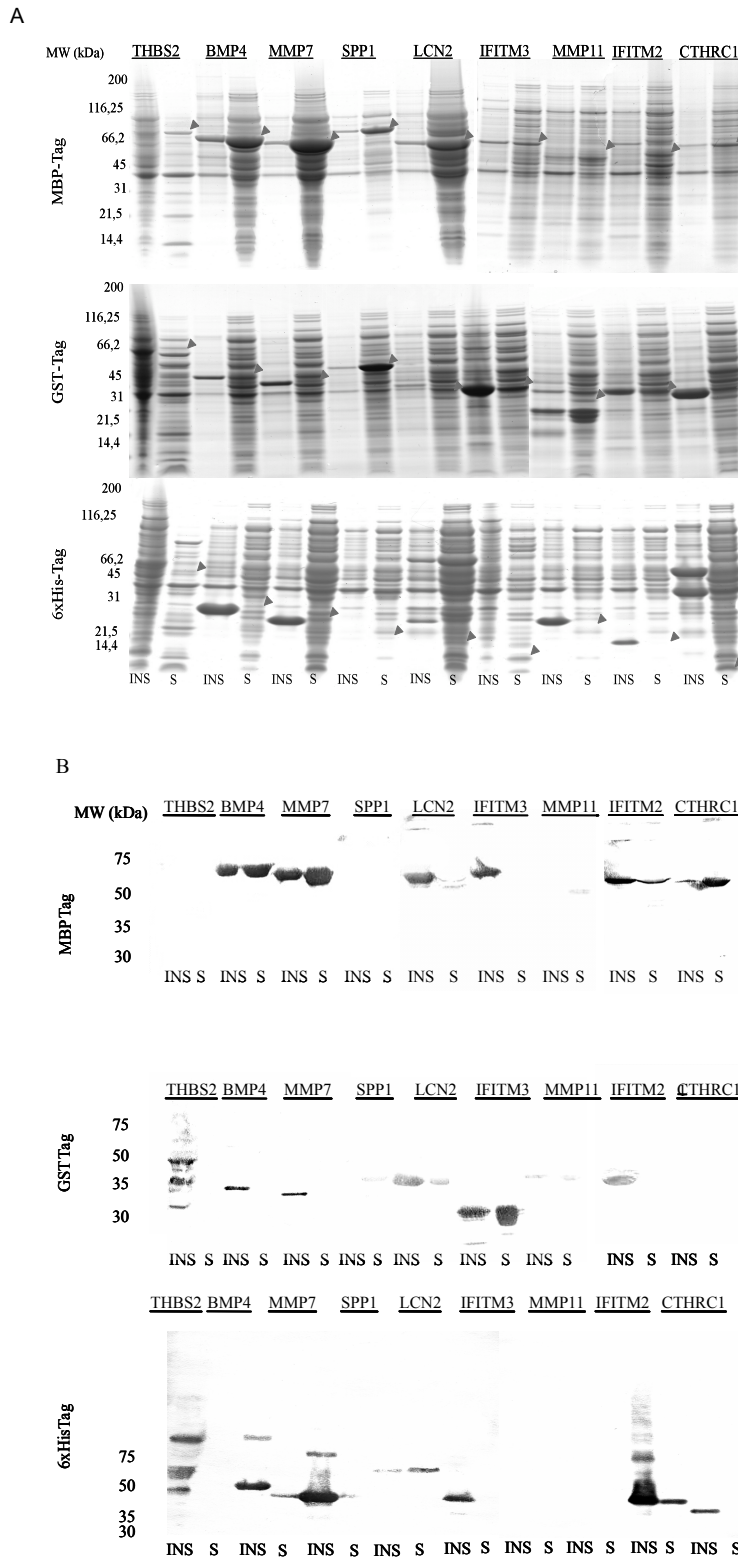


Figura 15: Expresión y solubilidad de las proteínas con secuencia parcial fusionadas a MBP, GST y 6xHis. Los cDNA se clonaron en tres vectores de expresión mediante recombinación homóloga y se transformaron células BL21(DE3). El crecimiento e inducción de la expresión de las bacterias se llevó a cabo como se describe en Materiales y métodos. Se cargaron muestras de las fracciones soluble e insoluble en geles de poliacrilamida SDS al 10%. A) Tinción con Azul brillante de Coomassie de los geles mostrando las fracciones soluble (S) e insoluble (INS). B) Análisis de la expresión mediante *Western blot* utilizando anticuerpos específicos frente a cada tag, indicado a la izquierda de cada imagen. El peso molecular en kDa se indica a la izquierda de cada imagen. Las cabezas de flecha indican el tamaño esperado de cada proteína recombinante.

Aunque ninguna proteína fue soluble con los tres *tags*, todas fueron correctamente expresadas con alguno de los tres *tags*. Por lo tanto, combinando los tres *tags* (MBP, GST y 6xHis) se obtuvieron todos los fragmentos proteicos seleccionados como productos solubles.

2.2 Expresión de proteínas a partir de clones de cDNA de secuencia completa

Para producir las proteínas completas, se obtuvieron clones de cDNA de secuencia completa a partir de la colección MGC para 24 de los 29 genes seleccionados (82%) y el resto por donaciones de otros laboratorios. Los cDNAs de los genes COL5A2, ITGA2 y SLC2A1 no se pudieron obtener. Las secuencias codificadoras, excepto el péptido señal, se amplificaron por PCR y se clonaron en pENTR/D-TOPO con una eficiencia variable entre un 15 y un 100% de las colonias analizadas. En tres casos, CA9, IGFBP3 e IFITM1, no se obtuvo el vector de entrada pENTR/D-TOPO. A continuación, los vectores de entrada se recombinaron con los tres vectores destino empleados anteriormente. Los resultados de la expresión y solubilidad de las proteínas se resumen en la Tabla 8.

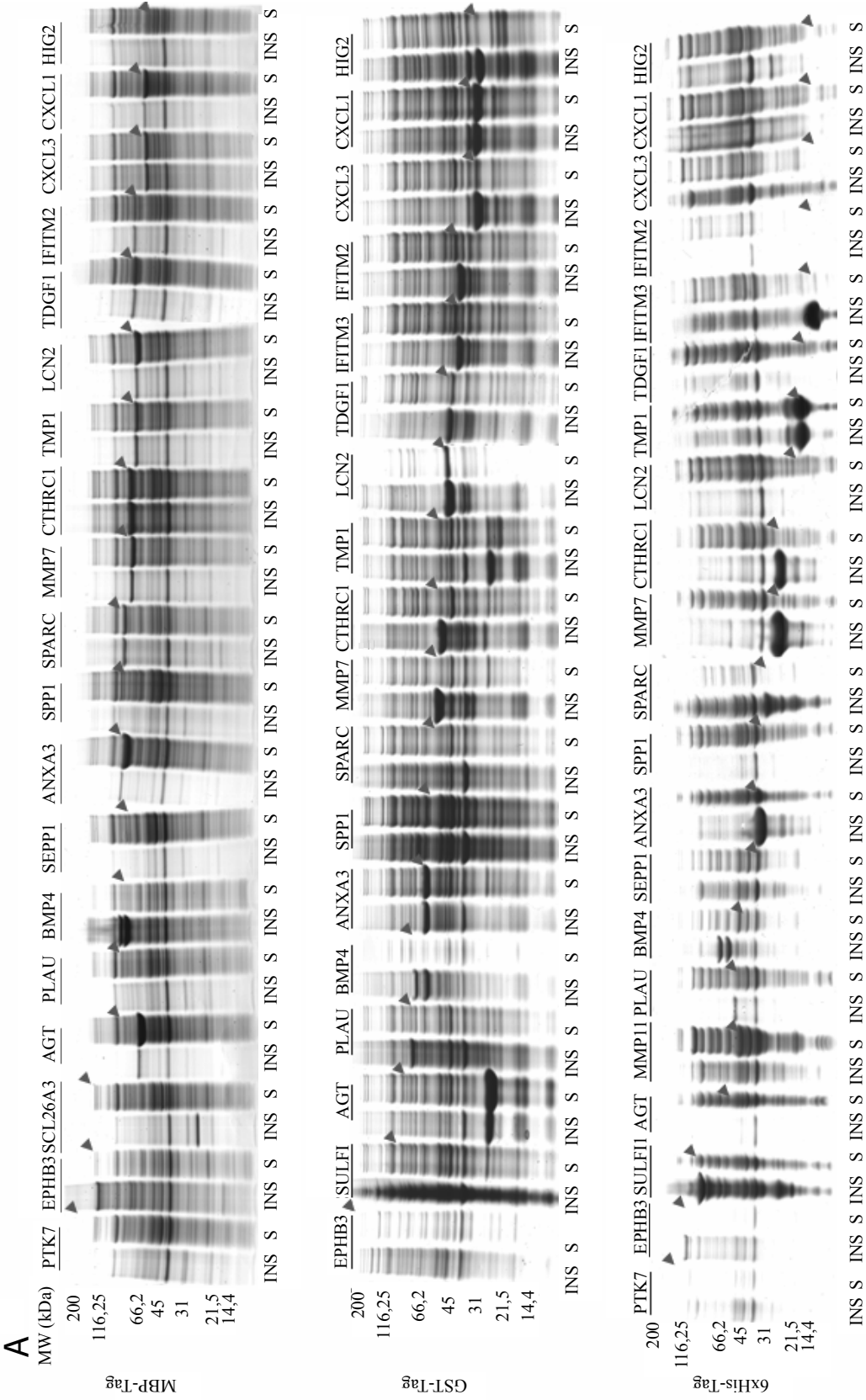
Con MBP se expresaron 13 de las 18 proteínas en pTH1 (72.2%); con GST se expresaron 15 de las 17 proteínas en pDEST15 (88%) y con 6xHis se expresaron 14 de las 21 proteínas en pDEST17 (66.7%). En relación a la solubilidad de las proteínas expresadas, se obtuvieron 11 proteínas como productos solubles en el caso de la fusión a MBP (84.6%), 9 en la fusión a GST (60%) y 8 en el caso de la fusión a 6xHis (57%). Todas las proteínas con un tamaño superior a 50 kDa aparecen en la fracción insoluble, excepto SULF1 con el *tag* 6xHis, independientemente del *tag* utilizado para la expresión en *E. coli* (Figura 16) (Madoz-Gúrpide y López-Serra *et al*, 2006)

En general, la obtención de vectores destino con clones de cDNA mayores de 1400 pb fue problemática, sugiriendo que la dificultad en la recombinación es mayor a medida que el tamaño del DNA a clonar es mayor. De hecho, la eficiencia de la clonación de cDNA con el sistema *Gateway* cuando el producto a clonar es menor o igual a 2 Kb es del 90% pero decrece considerablemente cuando los insertos son mayores (Marsischky y LaBaer, 2004).

Tabla 8. Niveles de expresión y solubilidad de los genes seleccionados clonados con tres *tag* diferentes. Los niveles de expresión se valoran de 0 a 3. 0: no hay banda visible en geles SDS-PAGE; 1: banda visible; 2: banda de tamaño medio; 3: banda de gran tamaño; *: tamaño de la banda observada incorrecto. Los niveles de solubilidad se valoran de 0 a 3. 0: no hay expresión en la fracción soluble; 1: banda minoritaria en la fracción soluble; 2: 50% de expresión en la fracción soluble; 3: mayor parte de la proteína en la fracción soluble.

Símbolo	Ontología	Total/Parcial	Peso molecular Da	Tag			Espectrometría de masas
				MBP E/S	GST E/S	His E/S	
PTK7	Transducción de señales	T	118.391	0/0	N	0/0	
		P	NC				
EPHB3	Angiogénesis	T	106.520	2/0	1/0	1/0	Sí
		P	NC				
SULF1	Apoptosis	T	101.026	N	1/0	3/1	Sí
		P	NC				
SLC26A3	Transportador de sulfato	T	84.671	0/0	N	N	
		P	NC				
AGT	Presión sanguínea	T	53.154	*	*	1/1	
		P	NC				
CA9	Factor de crecimiento	T	49.698	N	N	N	
		P	15.653	N	N	N	
PLAU	Factor de crecimiento	T	48.525	1/1	2/1	3/0	Sí
		P	NC				
THBS2	Angiogénesis	T	NC				
		P	47.898	0/0	2/2	0/0	Sí
BMP4	Factor de crecimiento	T	44.048	3/0	2/1	1/0	Sí
		P	23.867	3/3	3/2	3/0	Sí
SEPP1	Matriz extracelular	T	42.705	0/0	N	3/2	Sí
		P	NC				
ANXA3	Unión a calcio	T	36.375	2/3	3/2	3/1	Sí
		P	NC				
SPP1	Factor de crecimiento	T	35.422	0/0	1/0		Sí
		P	21.523	3/3	3/3	0/0	Sí
SPARC	Factor de crecimiento	T	34.632	2/2	0/0	2/1	Sí
		P	NC				
IGFBP3	Factor de crecimiento	T	31.674	N	N	N	
		P	NC				
MMP7	Matriz extracelular	T	29.677	3/3	3/0	3/0	Sí
		P	19.429	3/3	3/0	*	Sí
CTHRC1	Matriz extracelular	T	26.224	3/3	3/1	3/0	Sí
		P	11.813	2/3	3/0	1/0	Sí
MMP11	Matriz extracelular	T	23.628	N	N	2/3	
		P	13.956	2/3	2/2	*	Sí
TIMP1	Factor de crecimiento	T	23.171	2/2	0/0	0/0	Sí
		P	NC				
LCN2	Oncogén	T	22.588	3/3	3/2	3/2	Sí
		P	15.488	2/3	2/1	1/3	Sí
TDGF1	Factor de crecimiento	T	21.168	2/3	3/1	2/0	Sí
		P	NC				
IFITM3	Ciclo celular	T	14.632	N	3/1	3/1	Sí
		P	14.148	2/2	3/1	0/0	
IFITM2	Ciclo celular	T	14.546	1/1	3/0	0/0	Sí
		P	12.452	2/1	3/2	3/2	
IFITM1	Ciclo celular	T	13.964	N	N	N	
		P	13.364	N	N	N	
CXCL3	Factor de crecimiento	T	11.342	2/2	3/1	0/0	Sí
		P	NC				
CXCL1	Factor de crecimiento	T	7.864	3/3	3/2	0/0	Sí
		P	NC				
HIG2	Marcador de hipoxia	T	6.95	0/0	3/0	0/0	Sí
		P	NC				

NC: No hay clon disponible; T: cDNA de secuencia completa; P: cDNA de secuencia parcial; E: Nivel de expresión; S: Nivel de solubilidad; espacio en blanco: clonación en vectores *Gateway* fallida.



B

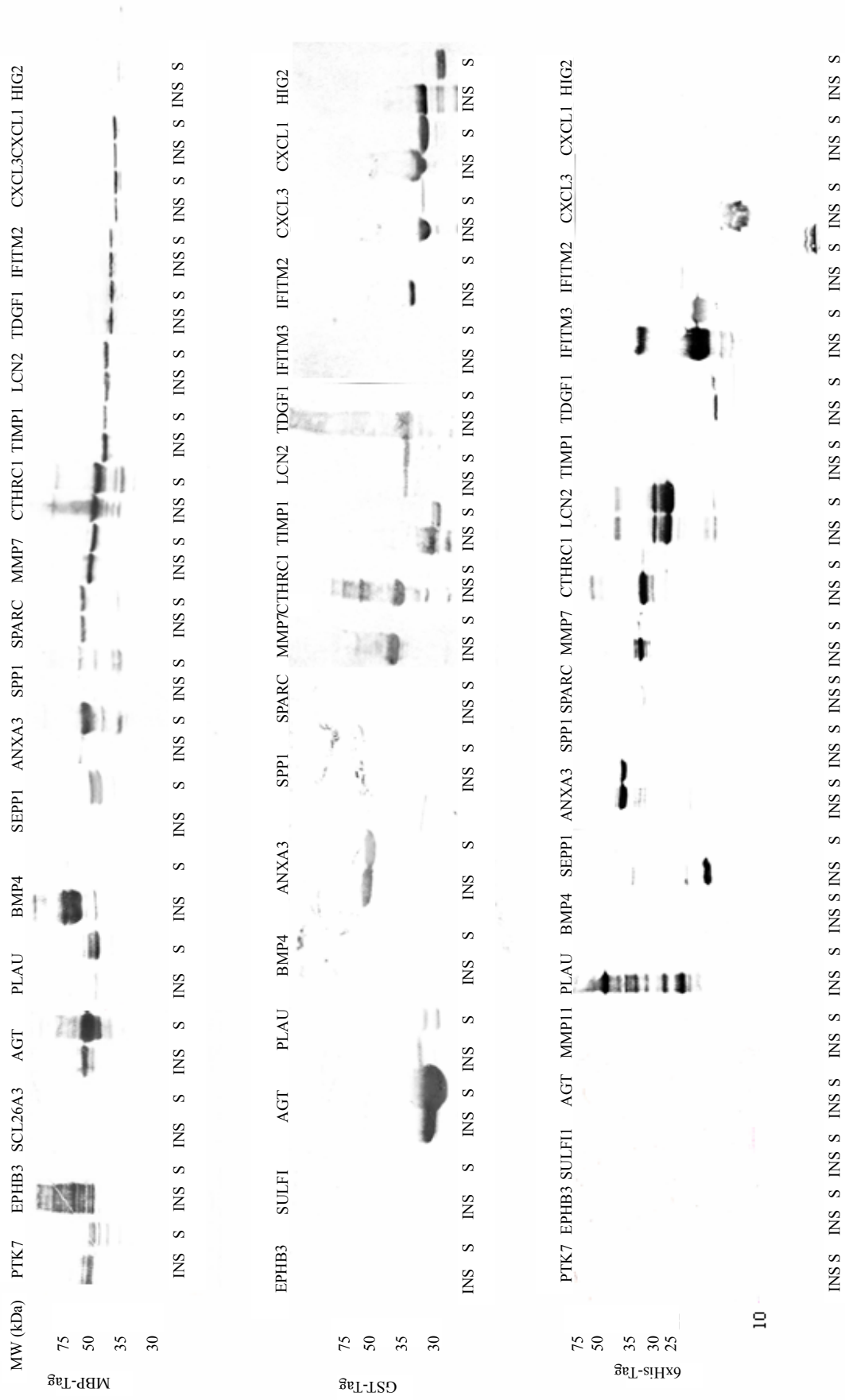


Figura 16. Expresión y solubilidad de las proteínas con secuencia completa fusionados a MBP, GST, 6xHis. La metodología experimental es la misma que en el caso de los clones de secuencia parcial. A) Tinción con Azul brillante de *Coomassie* de los geles mostrando las fracciones soluble (S) e insoluble (I). B) Análisis de la expresión mediante *western blot* utilizando anticuerpos específicos frente a cada *tag*. El peso molecular en kDa se indica a la izquierda de cada imagen. Las cabezas de flecha indican el tamaño esperado de da cada proteína recombinante.

2.3 Purificación de proteínas recombinantes

Las proteínas fueron purificadas para la producción de anticuerpos tanto monoclonales en ratón como policlonales en conejo y recombinantes utilizando la tecnología del despliegue en fagos. Todas las proteínas se purificaron mediante cromatografía de afinidad. Los resultados del proceso de purificación se muestran en la Figura 17. Las proteínas solubles fusionadas a MBP se purificaron utilizando una resina de amilasa. Una característica de todas las purificaciones con este *tag* es la presencia de bandas de menor tamaño al esperado que corresponden, probablemente, a fragmentos de degradación de la MBP. Para las proteínas de fusión a GST se obtuvieron resultados parecidos ya que en los geles de polipoliacrilamida-SDS se vieron bandas de menor tamaño al esperado. Respecto a las proteínas fusionadas al *tag* de histidinas, dada su insolubilidad, cuatro de ellas se purificaron en condiciones desnaturizantes utilizando un método de renaturalización en columna basado en la desnaturalización de las proteínas para su captura en las columnas de afinidad renaturalizándolas una vez fijadas a la columna al disminuir paulatinamente las concentraciones de urea y cloruro de guanidinio.

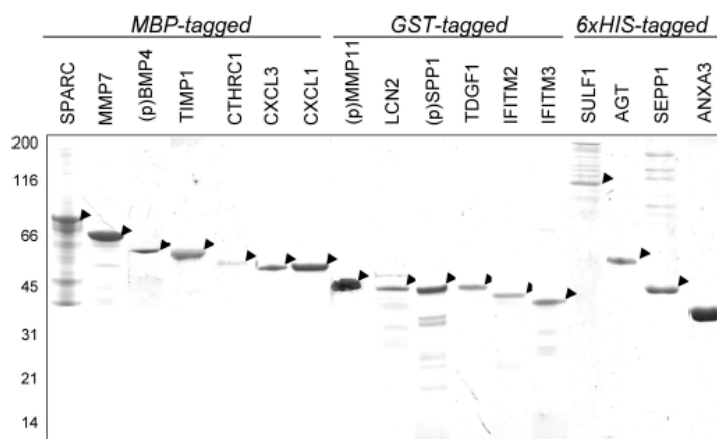


Figura 17. Purificación de las proteínas recombinantes mediante cromatografía de afinidad. Tinción con Azul brillante de *Coomassie* de geles SDS-PAGE donde se muestran las proteínas purificadas. Las fusiones a MBP se purificaron con resina de amilasa, las fusiones a GST se purificaron con columnas de glutatión y las fusiones a 6xHis se purificaron con columnas de Ni^{2+} . Los marcadores de peso molecular se muestran a la izquierda de la imagen. Las cabezas de flecha indican el tamaño esperado de da cada proteína recombinante. Las proteínas de secuencia parcial se identifican con una (p) antes del nombre.

Aunque la presencia de productos de degradación puede ser problemática para algunos usos de las proteínas (como su cristalización) no supone un problema importante en la generación de anticuerpos ya que todos los fragmentos derivan de la proteína de fusión que se utilizará como antígeno.

Cada proteína de fusión, ya sea en su forma completa o parcial, con los tres *tags* utilizados se analizó mediante espectrometría de masas MALDI-TOF para confirmar su identidad. Todas las proteínas de tamaño completo fueron correctamente identificadas excepto SCL26A3. En el caso de las proteínas de tamaño parcial las 9 fueron validadas correctamente.

3. Obtención de anticuerpos para la validación de las dianas seleccionadas

El siguiente paso en el proceso de validación requirió la preparación de anticuerpos específicos frente a las dianas genómicas. Se utilizaron varias estrategias experimentales diferentes. Por un lado, estrategias combinatorias, teóricamente más rápidas y de alto rendimiento, basadas en librerías de anticuerpos recombinantes expresados en la superficie del fago M13 (*Phage display*). Por otro lado, la producción de anticuerpos monoespecíficos frente a fragmentos de la proteína completa (método *PrEST*) en aquellos casos en que no se pudo obtener la proteína completa. Por último, las vías clásicas de producción de anticuerpos policlonales en conejo y anticuerpos monoclonales en ratón. En la Figura 18 se muestra un esquema de las diferentes estrategias utilizadas.

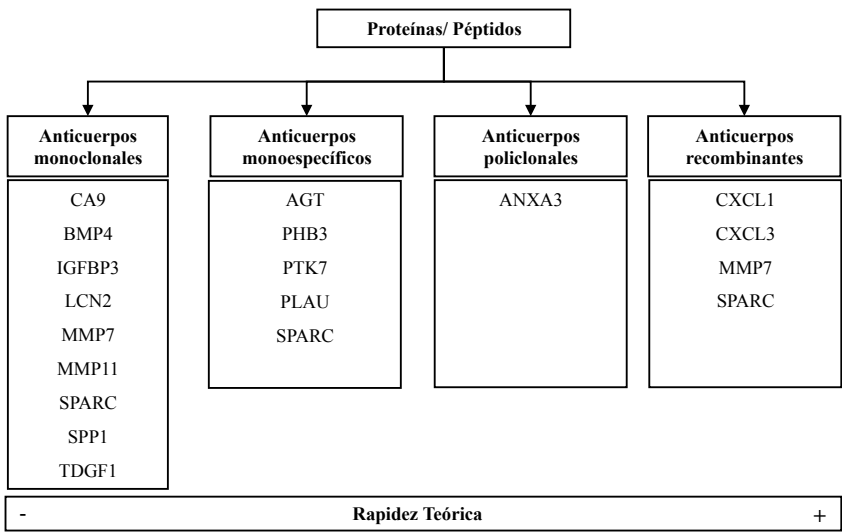


Figura 18. Estrategias utilizadas para la obtención de anticuerpos frente a las proteínas seleccionadas

3.1 Obtención de anticuerpos recombinantes humanos mediante despliegue en fagos (*Phage display*)

Para seleccionar anticuerpos recombinantes mediante la tecnología del despliegue en fagos se utilizaron dos librerías diferentes de anticuerpos humanos recombinantes, las librerías Tomlinson y las librerías Mehta, dado que uno de nuestros objetivos en la obtención de posibles anticuerpos terapéuticos. En la Figura 19 se muestra el esquema del proceso de obtención de anticuerpos utilizando esta técnica. Brevemente, las librerías de fagos recombinantes que presentan las scFv en su superficie se enfrentan a la proteína de interés. Las partículas virales que no se unan a la proteína diana serán eliminadas mediante los lavados mientras que los fagos unidos, que han reconocido el antígeno, se eluyen para infectar células de *E. coli* y amplificar los fagos específicos para un nuevo ciclo de selección. Tras tres ciclos de selección y enriquecimiento de fagos específicos, se pican colonias individuales para obtener scFv monoclonales que serán caracterizadas para determinar el número de secuencias únicas obtenidas y su reactividad frente al antígeno. Para la obtención de las scFv solubles se infectó con los fagos seleccionados una cepa de *E. coli* no-supresora, que reconoce el codon *stop* que separa las scFv de la pIII del fago.

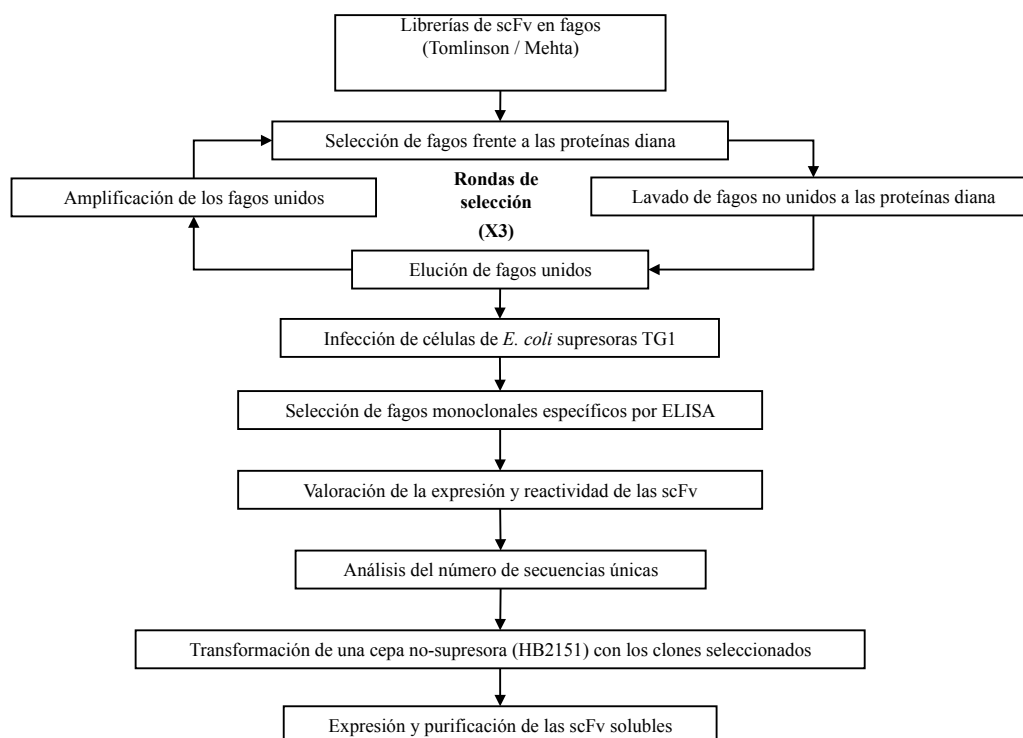


Figura 19. Esquema del proceso de selección de anticuerpos recombinantes mediante la tecnología del despliegue en fagos.

Como antígenos se utilizaron cuatro proteínas recombinantes de secuencia completa: CXCL3, MMP7 y SPARC fusionadas a MBP y CXCL1 fusionada a GST. Estas cuatro proteínas fueron seleccionaron por su implicación en la progresión neoplásica dado que todas ellas están involucradas en la invasión tumoral del tejido adyacente y la metástasis a otros tejidos.

3.1.1 Selección de scFv frente a las proteínas CXCL3, MMP7 y SPARC a partir de las librerías Tomlinson

Como primera aproximación para la obtención de scFv frente a las proteínas CXCL3, MMP7 y SPARC se utilizaron las librerías semisintéticas Tomlinson I+J. Se realizaron tres rondas de selección de fagos policlonales utilizando como antígenos las proteínas recombinantes fijadas en pocillos de placas de 96 pocillos. En la Figura 20 se observa el enriquecimiento en la especificidad de los fagos a lo largo del proceso determinado mediante ELISA de fagos policlonales tras cada ronda de selección, tanto frente a las proteínas de interés como al *tag* purificado. En el caso de CXCL3 se observa una fuerte reactividad cruzada con MBP, que se podría explicar en base al pequeño tamaño de la proteína CXCL3 (11,3 kDa) frente a MBP (42 kDa). En el caso de MMP7 y SPARC se observa un claro enriquecimiento específico frente a las correspondientes proteínas en comparación a la reactividad con MBP. En estos dos casos la relación de tamaños entre las proteínas y el *tag* es más similar, 29,68 kDa MMP7 y 34,6 kDa SPARC.

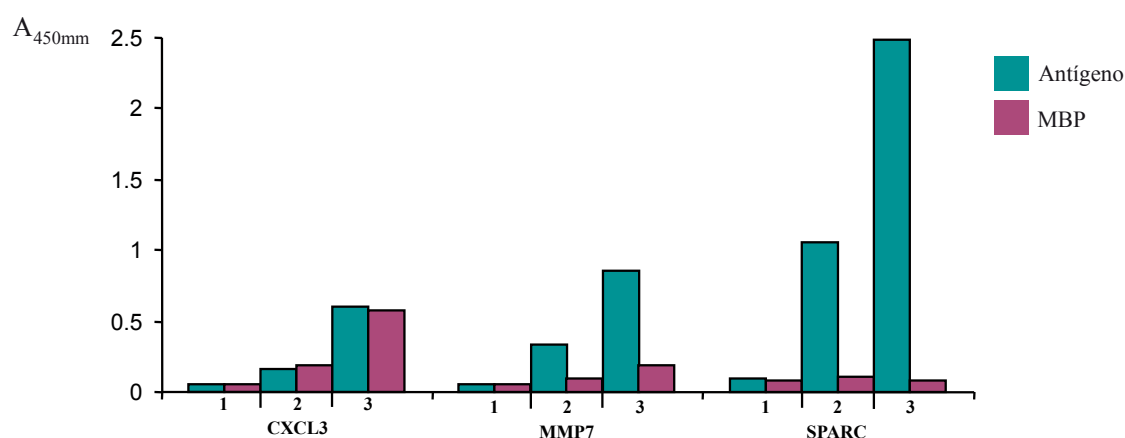


Figura 20. Titulación de fagos policlonales de las librerías Tomlinson. Se valoraron por ELISA los fagos policlonales de cada ronda de selección enfrentándolos tanto a los antígenos recombinantes purificados como a MBP. 1: Fagos de la primera ronda de selección; 2: segunda ronda de selección; 3: tercera ronda de selección.

Para obtener fagos monoclonales específicos frente a cada proteína diana se picaron colonias individuales a partir de la población de fagos policlonales eluidos en la tercera ronda de selección. Para CXCL3 se picaron 672 colonias individuales de las que se obtuvieron 60 fagos positivos por ELISA (8,9%). En el caso de MMP7 se picaron un total de 482 colonias individuales y se obtuvieron 329 fagos positivos por ELISA (68,3%). Para SPARC se picaron 384 colonias individuales y se obtuvieron 326 fagos positivos por ELISA (84,9%).

3.1.1.1 Reactividad de las scFv seleccionadas

Las scFv monoclonales se expresaron a partir de las colonias que resultaron positivas por ELISA para analizar su solubilidad y niveles de expresión. Las scFv solubles contienen la proteína pIII fusionada porque la cepa de *E. coli* utilizada en las rondas de selección de fagos, TG1, es una cepa supresora del codon *stop* TAG e introduce un residuo de glutamina en lugar del codon *stop*. El ELISA se llevó a cabo con la fracción periplásmica, que contiene las scFv debido a la presencia de la secuencia pelB en el fagémido.

De los 60 fagos positivos frente a CXCL3 se obtuvieron 11 scFv solubles y positivas por ELISA (18,3%). En el caso de MMP7, de los 329 fagos positivos se obtuvieron 9 scFv reactivas (2,7%) y para SPARC se obtuvieron 23 scFv positivas de los 326 fagos seleccionados (7%). Como se esperaba, el paso de fago a scFv disminuye significativamente el número de scFv solubles y funcionales. El reconocimiento del antígeno por parte de las scFv varía si están en la superficie del fago o en forma soluble.

3.1.1.2 Secuenciación y caracterización de las scFv seleccionadas

A continuación se analizó el número de secuencias únicas de scFv para cada antígeno. Las scFv funcionales se secuenciaron y las secuencias obtenidas se alinearon utilizando los programas ClustalW, que alinea múltiples secuencias simultáneamente, y GeneDoc, que muestra gráficamente los residuos compartidos por las secuencias analizadas. Se obtuvieron sólo 2 secuencias únicas de 11 (18%) frente a CXCL3: F6 y F8, 2 de 9 (22%) frente a MMP7: F7 y G11 y 4 de 23 (17%) frente a SPARC: A4, A5, A10 y B12. Como se muestra en la Figura 21, en todos los clones se observaron codones *stop* en la región hipervariable determinante de complementariedad 2 (CDR2) de la V_H. La presencia de estos codones *stop* impide expresar las scFv sin proteína pIII en cepas de *E. coli* no-supresoras como HB2151.

CXCL3

F6 : DRQLACKFYFKETVIMKYLLPTAAAGLLLLAAQPAMAEVOLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSTIAKR : 91
F8 : DRQLACKFYFKETVIMKYLLPTAAAGLLLLAAQPAMAEVOLLESGGGLV-PGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSTITVIG : 90

F6 : G-SRTRYADSVKGRFTISRDNSENKNTLYLQMNLSRAEDTAVYYCAKRIATFDYWGQGLTVTVSSGGGGSGGGSGGGGSTDIQMTQSPSSL : 181
F8 : GTRTLYADSVKGRFTISRDNSENKNTLYLQMNLSRAEDTAVYYCAKPARGFEDYWGQGLTVTVSSGGGGSGGGSGGGGSTDIQMTQSPSSL : 181

F6 : SVGDRVTITCRASQSISSYLNWYQKPGKAPKLLIYAASRLQSGVPSRFSGSGSGTDFTLTISLQPEDFATYYCQQAAYEPFTFGGQGTKV : 272
F8 : SVGDRVTITCRASQSISSYLNWYQKPGKAPKLLIYAASRLQSGVPSRFSGSGSGTDFTLTISLQPEDFATYYCQQAAYEPFTFGGQGTKV : 272

F6 : EIKRAAAHHHHHH : 285
F8 : EIKRAAAHHHHHH : 285

MMP7

F7 : DRQLACKFYFKETVIMKYLLPTAAAGLLLLAAQPAMAEVOLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSTIAK : 90
G11 : DRQLACKFYFKETVIMKYLLPTAAAGLLLLAAQPAMAEVOLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSSITS : 90

F7 : RP-KRYADSVKGRFTISRDNSENKNTLYLQMNLSRAEDTAVYYCAKRRSLGFDYWGQGLTVTVSSGGGGSGGGSGGGGSTDIQMTQSPSSL : 179
G11 : PGTKRYADSVKGRFTISRDNSENKNTLYLQMNLSRAEDTAVYYCAKPGRRHFDYWGQGLTVTVSSGGGGSGGGSGGGGSTDIQMTQSPSSL : 179

F7 : SASVGDRVTITCRASQSISSYLNWYQKPGKAPKLLIYNASRLQSGVPSRFSGSGSGTDFTLTISLQPEDFATYYCQQRIRYPTTFGQG : 269
G11 : SASVGDRVTITCRASQSISSYLNWYQKPGKAPKLLIYNASRLQSGVPSRFSGSGSGTDFTLTISLQPEDFATYYCQQRIRYPTTFGQG : 269

F7 : TKVEIKRAAAHHHHHH : 285
G11 : TKVEIKRAAAHHHHHH : 285

SPARC

A4 : DRQLACKFYFKETVIMKYLLPTAAAGLLLLAAQPAMAEVOLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSAIR : 89
A10 : DRQLACKFYFKETVIMKYLLPTAAAGLLLLAAQPAMAEVOLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSGISK : 90
B12 : DRQLACKFYFKETVIMKYLLPTAAAGLLLLAAQPAMAEVOLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSTIAK : 90
A5 : DRQLACKFYFKETVIMKYLLPTAAAGLLLLAAQPAMAEVOLLESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVTIT-R : 89

A4 : SGKMTRYADSVKGRFTISRDNSENKNTLYLQMNLSRAEDTAVYYCAKSVRRFDYWGQGLTVTVSSGGGGSGGGSGGGGSTDIQMTQSPSSL : 179
A10 : SGA-TRYADSVKGRFTISRDNSENKNTLYLQMNLSRAEDTAVYYCAKRIIRFDYWGQGLTVTVSSGGGGSGGGSGGGGSTDIQMTQSPSSL : 179
B12 : RG-KRYADSVKGRFTISRDNSENKNTLYLQMNLSRAEDTAVYYCAKRRSLGFDYWGQGLTVTVSSGGGGSGGGSGGGGSTDIQMTQSPSSL : 179
A5 : GGAT-TRYADSVKGRFTISRDNSENKNTLYLQMNLSRAEDTAVYYCAKPTAKFDYWGQGLTVTVSSGGGGSGGGSGGGGSTDIQMTQSPSSL : 178

A4 : SASVGDRVTITCRASQSISSYLNWYQKPGKAPKLLIYGASSLQSGVPSRFSGSGSGTDFTLTISLQPEDFATYYCQQHNNFPHITFGQG : 269
A10 : SASVGDRVTITCRASQSISSYLNWYQKPGKAPKLLIYSASGLQSGVPSRFSGSGSGTDFTLTISLQPEDFATYYCQQRHLPITFGQG : 269
B12 : SASVGDRVTITCRASQSISSYLNWYQKPGKAPKLLIYKASLLQSGVPSRFSGSGSGTDFTLTISLQPEDFATYYCQQEWRRGPTFGQG : 269
A5 : SASVGDRVTITCRASQSISSYLNWYQKPGKAPKLLIYSASVLSQSGVPSRFSGSGSGTDFTLTISLQPEDFATYYCQQAHSAPITFGQG : 268

A4 : TKVEIKRAAAHHHHHH : 285
A10 : TKVEIKRAAAHHHHHH : 285
B12 : TKVEIKRAAAHHHHHH : 285
A5 : TKVEIKRAAAHHHHHH : 284

Figura 21. Secuencias de la V_H y V_L de las scFvs seleccionadas a partir de las librerías Tomlinson. Los puntos negros representan los codones *stop* TAG presentes en la secuencia, casi todos ellos en el CDR2 de la V_H.

En la Tabla 9 se resume el número de fagos, scFv y secuencias únicas obtenidas para cada proteína, en ella se puede observar la relativa ineficiencia del proceso de producción de scFv a partir de estas librerías. De las 2 secuencias únicas seleccionadas frente a CXCL3 solamente F6 reaccionó por *western blot* frente a la proteína recombinante y no frente a MBP. En el caso MMP7 sólo G11 es positivo por *western blot*. En el caso de SPARC resultaron positivos 2 de los 4 clones testados, A4 y B12 (Figura 22). Por tanto, para los estudios posteriores se seleccionaron F6, G11, A4 y B12.

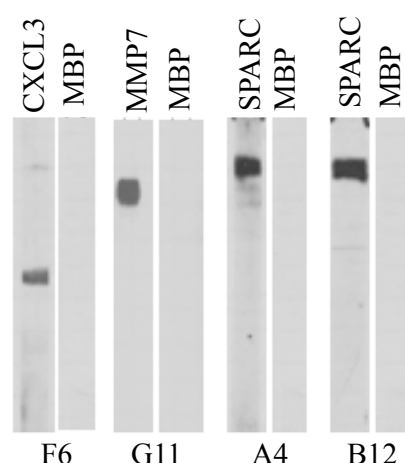


Figura 22. Reactividad de las scFv seleccionadas por *Western blot* frente a sus antígenos. Se transfirieron las proteínas purificadas a membranas de nitrocelulosa y se incubaron con las respectivas scFv. Como anticuerpo secundario se utilizó anti c-Myc conjugado a peroxidasa y se reveló con ECL.

3.1.1.3 Preparación de scFv solubles

La presencia de codones *stop* en todos los clones seleccionados impidió la expresión de las scFv como productos solubles sin pIII en una cepa de *E. coli* no-supresora. Para solucionar este problema fue necesario mutar el codon TAG presente en la unión entre las scFv y la pIII por el codon *stop* TAA, que sí es reconocido como *stop* por las cepas supresoras, como TG1, siguiendo el protocolo descrito por Barderas *et al* (2006). Con este método, tan sólo fue necesario mutar el codon que separa las scFv de la proteína pIII en lugar de realizar múltiples mutagénesis dirigidas para eliminar cada codon *stop* presente en las CDR.

Con este método se mutaron las scFv seleccionadas F6 (CXCL3), G11 (MMP7) y A4 y B12 (SPARC). En la Figura 23 se muestra el resultado de la expresión y del

western blot. Se observa que hubo sobreexpresión de las scFv en todos los clones excepto en el clon B12 antiSPARC aunque los niveles de scFv soluble obtenidos no son muy elevados en ningún caso a la vista de la tinción con Azul brillante de *Coomassie* de las fracciones periplásmicas.

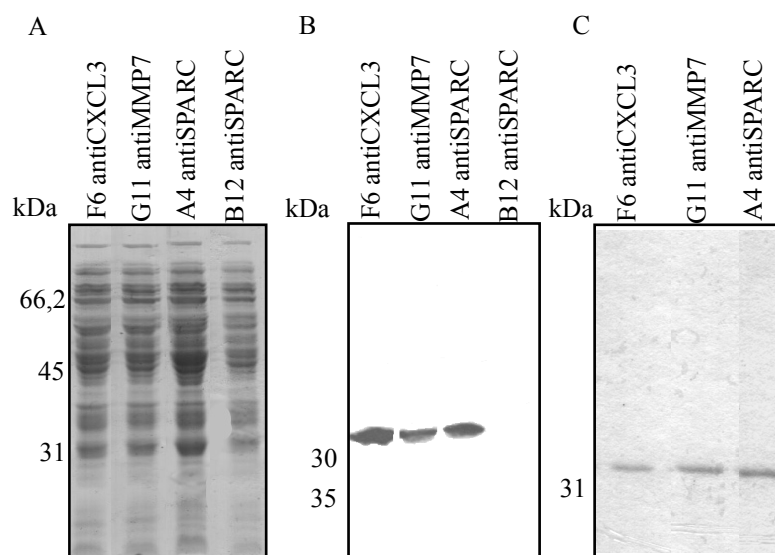


Figura 23. Expresión y purificación de las scFv tras mutar el codon *stop* TAG presente entre la scFv y la proteína pIII. A) Tinción con Azul brillante de *Coomassie* de las fracciones periplásmicas de TG1 tras llevar a cabo la expresión de las scFv sin pIII B) *Western blot* frente a las scFv en el periplasma, como anticuerpo se utilizó anti His conjugado a Peroxidasa. C) Tinción con *Coomassie blue* de las scFv purificadas con la resina Talon.

A pesar de los bajos niveles de expresión obtenidos, se purificaron las scFv utilizando cromatografía IMAC y columnas de Ni^{2+} . Como se observa en la figura 23C el grado de pureza de las scFv purificadas fue muy alto pero los niveles de proteína purificados continuaron siendo muy bajos ($<50 \text{ ng}/\mu\text{l}$). La llegada de las librerías Mehta al laboratorio y los problemas surgidos con las scFv seleccionadas a partir de las librerías Tomlinson nos hizo centrarnos en la obtención de nuevas scFv a partir de las librerías Mehta.

3.1.2 Selección de fagos a partir de la librerías Mehta

Dados los problemas que se encontraron con las librerías Tomlinson (presencia de codones *stop* y bajo rendimiento en la producción de scFv) decidimos utilizar las librerías *naïve* Mehta, que tienen un tamaño de $1,5 \times 10^{10}$ fagos, considerablemente mayor al de las librerías Tomlinson ($1,4 \times 10^8$).

Con estas librerías se seleccionaron scFv específicas frente a las mismas proteínas utilizadas previamente con las librerías Tomlinson, CXCL3, MMP7 y SPARC,

y CXCL1. El proceso de selección se repitió como se ha descrito previamente para las librerías Tomlinson. La población de fagos policlonales después de las sucesivas rondas de selección se valoró por ELISA (Figura 24). En todos los casos se produce un aumento de la señal específica frente a la proteína de interés en relación al *tag* a medida que transcurren las rondas de selección. Se observa que los valores de absorbancia obtenidos con estas librerías son superiores a los obtenidos con las librerías Tomlinson.

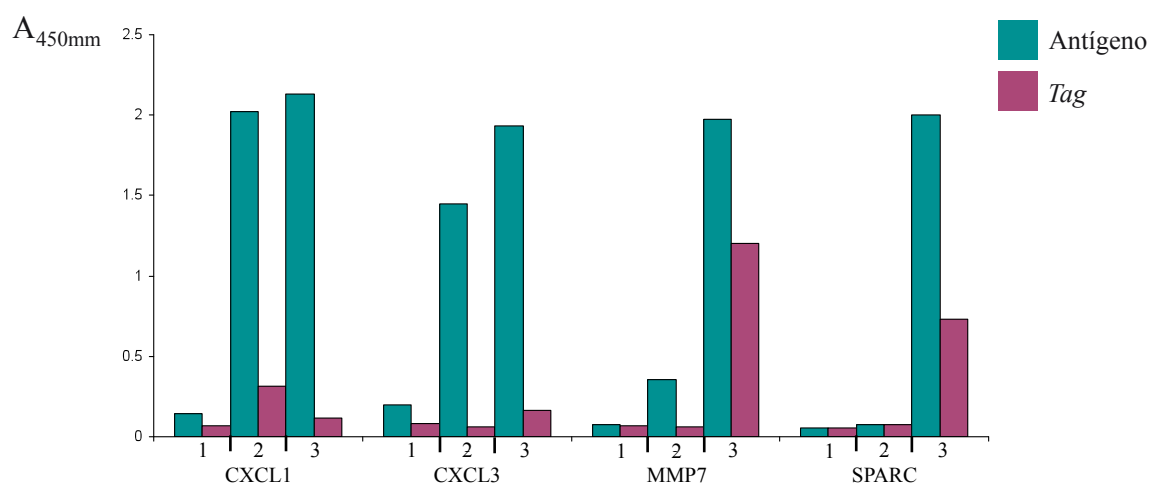


Figura 24. ELISA de fagos policlonales de la librería Mehta. Los fagos policlonales de cada ronda de selección y se enfrentaron tanto a los antígenos recombinantes purificados como a MBP y GST como control de la especificidad del proceso. Debajo de cada serie se especifica la proteína utilizada como antígeno; 1, 2 y 3 son los fagos eluidos de la primera, segunda y tercera ronda de selección.

Se picaron colonias individuales a partir de la tercera ronda de selección de fagos para analizar la especificidad de los fagos monoclonales. Se picaron un total de 240 colonias de la tercera ronda de selección frente a CXCL3 de las que resultaron positivas 30 (12,5%). Para MMP7, de 192 colonias se obtuvieron 16 positivas (8,3%) y en el caso de SPARC se obtuvieron 41 casos positivos de 192 colonias (21,3%); por último, para CXCL1, de 96 colonias se obtuvieron 66 positivas (68,75%). El alto porcentaje de fagos positivos obtenido frente a CXCL1, fusionada a GST, en relación al resto de proteínas, fusionadas a MBP, se debe probablemente el menor peso molecular de GST (25 kDa) en comparación a MBP (42 kDa), al permitir una mayor exposición de la proteína a la que está fusionado. Por otro lado, también se observa que los porcentajes de fagos positivos son mayores que los obtenidos con las librerías Tomlinson.

3.1.2.1 Especificidad de las scFv seleccionadas

Una vez seleccionados los fagos monoclonales específicos frente a las proteínas diana, se testó la funcionalidad de las scFv que codifican. Para esto se hizo un ELISA de scFv monoclonales utilizando la fracción periplásmica y anti c-myc como secundario. En el caso de CXCL3, de las 30 colonias analizadas se obtuvieron 8 scFv funcionales (26,6%). Para MMP7 se obtuvieron 6 scFv funcionales de las 16 colonias analizadas (37,5%). En el caso de SPARC se analizaron 41 scFv de las que resultaron funcionales 37 (90,2%). Por último, en el caso de CXCL1 se obtuvieron 20 scFv funcionales de un total de 66 analizadas (30,3%).

3.1.2.2 Secuenciación de las scFv seleccionadas

En la Tabla 9 se resume el número de fagos, scFv y secuencias únicas obtenidas para cada proteína. Tras el análisis se comprobó que 3 de las 8 secuencias analizadas en el caso de CXCL3 eran únicas, A1, E12 y F9. En el caso de MMP7 se obtuvieron 2 scFv de secuencia única de las 6 analizadas, B2 y F8. Para SPARC se comprobó que tan sólo había 1 secuencia única, A3, entre las 37 analizadas. Para CXCL1 se obtuvieron 2 secuencias únicas, B6 y D4, de 20 analizadas. Las secuencias únicas para cada clon se muestran en la Figura 25.

Tabla 9. Resumen de la obtención de anticuerpos recombinantes humanos únicos a partir de las librerías y Mehta

Librerías	Proteína	Colonias picadas	fagos	scFv	secuencias únicas
Tomlinson	CXCL3	672	60	11	2
	MMP7	482	329	9	2
	SPARC	384	326	23	4
Mehta	CXCL3	240	30	8	3
	MMP7	192	16	6	2
	SPARC	192	41	37	1
	CXCL1	96	66	20	2

A la vista de los resultados del ELISA de fagos policlonales, en un primer momento pensamos que la influencia del *tag* era fundamental a la hora de obtener un mayor número de scFv específicos frente a la proteína diana; sin embargo, el número de scFv funcionales únicas fue bajo en todos los casos, independientemente del *tag* utilizado, MBP o GST.

CXCL1

```

      *      20      *      40      *      60      *      80
B6 : PAMAQVQLVQSGGGVVPGRSLRLSCAASGFTFSSSYGMHWVRQAPGKGLQWLASISDGGYERYYPDSVRGQFTISRDN SKNTLYLQ : 86
D4 : PAMAQVQLVQSGGGLVQPGSLRLSCAASGFTFSSNHDHMHVVRQAPGKGLQWLASISNGQSTYYADSVKGRFTISRDN SKNTLYLQ : 86

      *      100     *      120     *      140     *      160     *
B6 : MNSLRGDDTAVYFCARESRNMDLDTWGQGTITVTVSSGGGGSGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGG : 172
D4 : MNSLRPEDTAVYFCARESSGWFYDYGQGTITVTVSSGGGGSGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGG : 170

      180      *      200      *      220      *      240      *      2
B6 : QWYQKPGQSPVITVIEDNCRPSGVDRFSGSIDSSNSASLTISGLKTEDEADYYCQSYDSNN-LWVFGGGTKLTVLGQPKAAP : 257
D4 : SWYVQKPGQSPVITVIEDNCRPSGVDRFSGSIDSSNSASLTISGLKTEDEADYYCQAWDTITGSIYVFGGGTKLTVLGQPKAAP : 254

      60
B6 : AAA : 260
D4 : AAA : 257

```

CXCL3

```

      *      20      *      40      *      60      *      80
A1 : PAMAQVQLVQSGGGVVPGRSLRLSCAASGFTFSSSYGMHWVRQAPGKGLEWVAVISYDGSNNY--YADSVKGRFTISRDN SKN : 81
F6 : PAMAQVQLVQSGGGLVQPGSLRLSCAASGFTFSSSYGMHWVRQAPGKGLEWVAVISYDGSNNY--YADSVKGRFTISRDN SKN : 81
E12 : PAMAQVQLVQSGGGLVQPGSLRLSCAASGFTFSSSYGMHWVRQAPGKGLEWVAVISYDGSNNY--YADSVKGRFTISRDN SKN : 84

      *      100     *      120     *      140     *      160     *
A1 : TLYLQMNSLRAEDTAVYFCARMGYYDYFFDYWGQGTITVTVSSGGGGSGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGG : 164
F6 : TLYLQMNSLRAEDTAVYFCARGY----HFDSWGQGTITVTVSSGGGGSGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGG : 162
E12 : QFSLQMNSLRAEDTAVYFCARLDLVG-LDAFDITWGQGTITVTVSSGGGGSGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGG : 167

      180      *      200      *      220      *      240      *
A1 : ---NIGSKSVHWYQOKPGQAPVIVVYDSDRFSGTFRFSGSN--SGNTATLTISVEAGDEADYYCQVMDRSGSAHPVFCG : 243
F6 : SLVHSDGNTILNWFQQRPGSPRRLLIKVSNRDSGVDRFSGSG--SGTDFTLKISVEAEADVGVVYCMQG--IYWPRTFCG : 243
E12 : G---SLASNLVQWYQQRPGSAPTTVIYEINCRPSGVDRFSGSIDSSNSASLTISGLKTEDEADYYCQSF--SNVIFGGG : 247

      260      *
A1 : VLVLGQPKAAPSA : 258
F6 : LEHK-RTVAAPTAA : 257
E12 : LTVLVSQPKAAPSA : 262

```

MMP7

```

      *      20      *      40      *      60      *      80
B2 : PAMAQVQLVQSGAEVKKPGSSSVKVSCTASGCTLSRVVIMHWVRQAPGORLEWVGWINAGNDTEYSQKFOGVTISRDTASTAYMEL : 87
F8 : PAMAQVQLVQSGGGLVQPGSLRLSCAASGFTFSSSYGMHWVRQAPGKGLEWVANIKEDGSEKYYVDSVGRFTISRDN AKNSLYLQ : 87

      *      100     *      120     *      140     *      160     *
B2 : SSLRSED TAVYFCARGRASCNNTSCYFNDGFDITWGQGTITVTVSSGGGGSGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGG : 174
F8 : NSLRSED TAVYFCATGRN-----AFGLWGQGTITVTVSSGGGGSGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGG : 163

      180      *      200      *      220      *      240      *      260
B2 : IGSNYVYVYQQLPGTAPKLLIYRNHHRPSGVDPDRFSGSK--SGTSDSLAISGLRSEDEADYYCAAWDDSLSGWVFGGGTKLTVLGQ : 259
F8 : IASFSVQWYQQRPGSAPSTVIEADNCRPSGVDPDRFSGSIDSSNSASLTISGLKTEDEADYYCQSYDR--SNVIFGGGTKLTVLGQ : 248

      B2 : KAPPSAAA : 267
      F8 : KAAPSA : 256

```

SPARC

```

      *      20      *      40      *      60      *      80
A3 : PAMAQVQLVQSGGGVVPGRSLRLSCAASGFTFSSSYGMHWVRQAPGKGLEWVAVISYDGSNNKYADSVKGRFTISRDN SKNTLYLQMN : 88

      *      100     *      120     *      140     *      160     *
A3 : SLRAEDTAVYFCARGFSEFLDYVYVYGMVWGQGTITVTVSSGGGGSGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSGGGG : 176

      180      *      200      *      220      *      240      *      260
A3 : SWYQKPGQAPVIVYVYGNRPSGVDRFSGSIDSSNSASLTISGLKTEDEADYYCQSYDR--SNVIFGGGTKLTVLGQPKAAPSA : 263

```

Figura 25. Secuencia de las scFv seleccionadas a partir de la librería Mehta. Tras secuenciar las scFv positivas para cada clon se alinearon las secuencias obtenidas mediante el programa ClustalW para determinar el número de secuencias únicas seleccionadas. En esta figura se muestran las secuencias de las scFv seleccionadas frente a CXCL1, CXCL3, MMP7 y SPARC.

Una de las ventajas de estas librerías con respecto a las Tomlinson es la ausencia de codones *stop* en la secuencia codificadora de las scFv por lo que el paso de la cepa de *E. coli* TG1 a la HB2151 fue inmediato. Se infectaron células de la cepa de *E. coli* no-supresora HB2151 con los fagos A1, F9 y E12 antiCXCL3, B6 y D4 antiCXCL1, B2 y F8 antiMMP7 y A3 antiSPARC. Tras inducir la expresión de las scFv con IPTG, los niveles de expresión en el periplasma de las scFv sin pIII fueron elevados. Por otro lado, se testó su reactividad frente al antígeno purificado por *western blot* comprobándose que las scFv reconocieron específicamente el antígeno frente al que fueron seleccionadas (Figura 26).

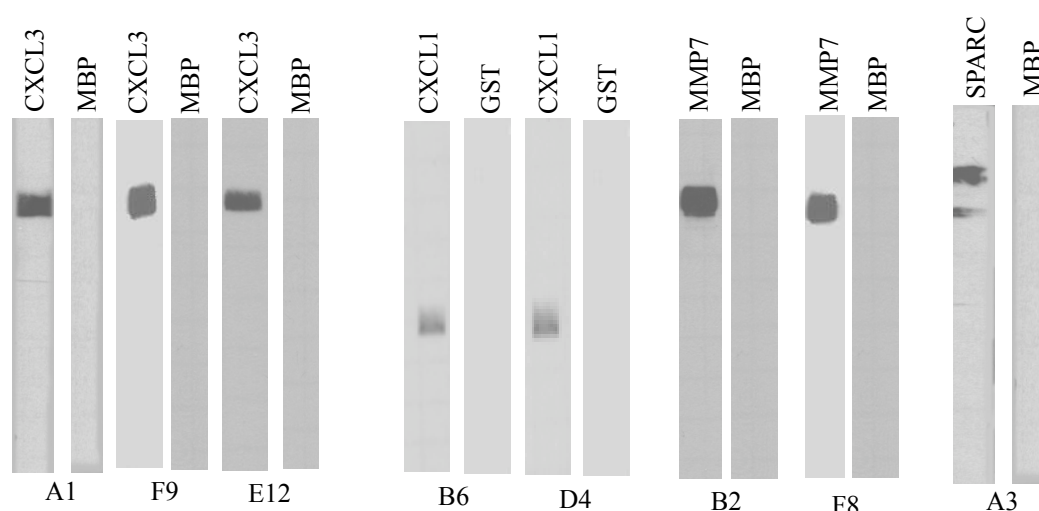


Figura 26. Western blot con las scFv seleccionadas frente a los antígenos purificados. Se transfirieron las proteínas purificadas y el *tag* a membranas de nitrocelulosa. En los dos casos analizados, la unión al *tag* MBP fue negativa mientras que las scFv sí reconocen la proteína frente a la que fueron seleccionadas. Como anticuerpo secundario se utilizó anti c-Myc y como terciario anti-Mouse conjugado a peroxidasa.

3.1.2.3 Expresión de las scFv Mehta conteniendo 6xHis

El fagémido en el que se construyeron las librerías Mehta, pFARBER, sólo contiene el *tag* c-myc, que no resulta adecuado para el análisis de la expresión de proteínas en cáncer colorrectal dado que deriva de la proteína humana c-myc, sobreexpresada en tumores que presenten alterada la vía APC. Para solucionar este inconveniente las scFv se expresaron en el vector pET28b, que tiene un *tag* 6xHis en el extremo carboxi-terminal. En la Figura 27A se observa que todas las scFv se expresaron en la fracción soluble y presentaron correctamente el *tag* 6xHis, detectado por *western blot* (Figura 27B).

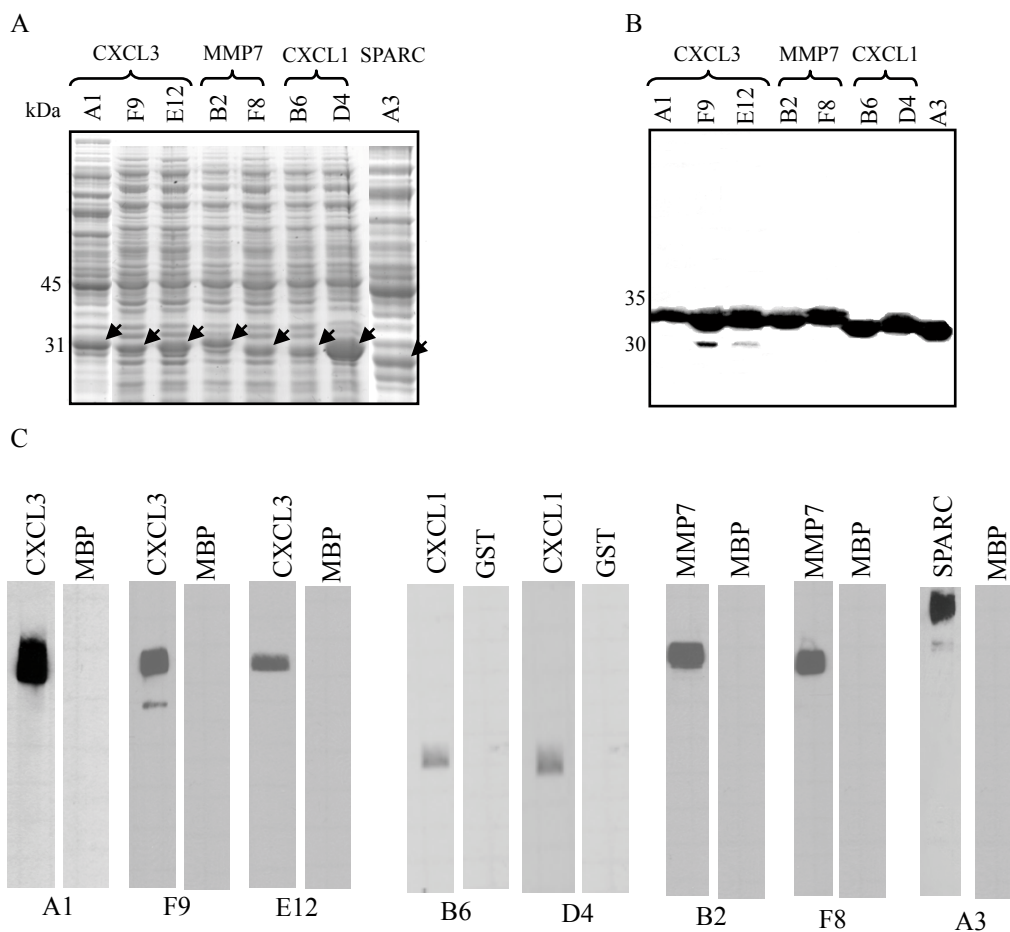


Figura 27. Expresión y solubilidad de las scFv Mehta en el vector pET28b. Se indujo la expresión de las scFv con IPTG tras transformar células BL21 (DE3) con las scFv subclonadas en el vector pET28b. A) Tinción con Azul brillante de *Coomassie* de las fracciones solubles de cada scFv. Las flechas señalan las bandas correspondientes a las scFv. B) *Western blot* de las fracciones solubles utilizando anti His conjugado a peroxidasa. C) Reconocimiento de los antígenos purificados por las scFv. Se transfirieron las proteínas purificadas a membranas de nitrocelulosa y se incubaron con las scFv correspondientes, como anticuerpo secundario se utilizó anti His conjugado a peroxidasa.

Se comprobó la capacidad de las scFv de reconocer al antígeno por *western blot* utilizando las fracciones solubles. En la Figura 27C se muestra que todos los clones fueron específicos frente a la proteína de interés y no frente al *tag*.

Las scFv se purificaron a partir de las fracciones solubles de cada clon utilizando columnas de Ni^{2+} . Se observaron dos picos de elución correspondientes a las scFv y al Factor de Elongación Tu de *E. coli* (analizado por huella peptídica en MALDI-TOF) y (Figuras 28A y B). En todos los casos se obtuvieron las scFv purificadas a una gran concentración (Figura 28C).

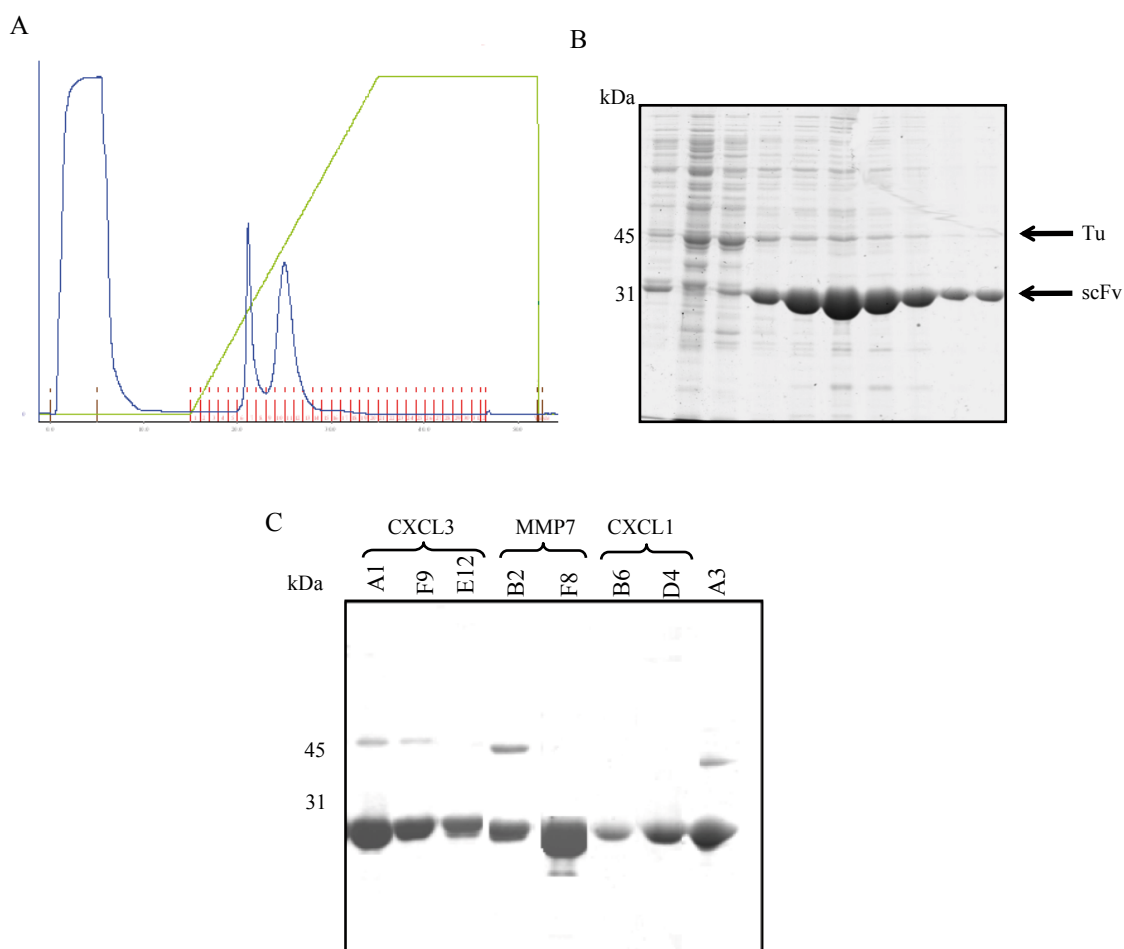


Figura 28. Purificación de las scFv Mehta mediante cromatografía de afinidad a metal inmovilizado (IMAC). Se utilizó una resina de Co^{2+} y un equipo ÄKTA Prime para llevar a cabo la purificación de las scFv en columnas de Ni^{2+} . A) Tinción con Azul brillante de *Coomassie* de las fracciones de elución de la resina de Co^{2+} . Como ejemplo se muestra la purificación de la scFv A3 antiSPARC. B) Cromatograma de la purificación de las scFv mediante cromatografía de afinidad en columna de Ni^{2+} . La línea azul muestra la absorbancia a 595 nm. La línea verde muestra el porcentaje de tampón de elución con imidazol en la columna. C) Tinción con Azul brillante de *Coomassie* de las fracciones de elución obtenidas en la purificación del scFv A3 antiSPARC. D) SDS-PAGE de las scFv seleccionadas purificadas.

3.1.2.4 Aplicación de las scFv al análisis de la expresión de proteínas diana en líneas celulares de cáncer colorrectal

Se analizaron los niveles de expresión de CXCL3 y SPARC en cuatro líneas de cáncer de colon, CaCo2, Colo205, Sw480 y Sw480, utilizando las scFv A1 anti CXCL3 y A3 anti SPARC. En ningún caso se observó sobreexpresión de las proteínas seleccionadas en las líneas celulares. Se descartó la inespecificidad de los anticuerpos ya que los controles positivo y negativo con las proteínas diana y tag MBP purificados funcionaron correctamente. Para estudiar si los anticuerpos frente al tag de histidinas no

eran suficientemente específicos se decidió subclonar las scFv en un vector con un *tag* 3xFLAG, de alta especificidad.

3.1.2.5 Expresión y purificación de las scFv seleccionadas con 3xFLAG

El vector pSANG10-3F contiene el *tag* 3xFLAG en el extremo carboxilo-terminal de las scFv. La secuencia del *tag* 3xFLAG se muestra en la Figura 29A. Este *tag* es muy apropiado a la hora de utilizar las scFv en inmunohistoquímica (Martin et al, 2006). Las scFv seleccionadas con las librerías Mehta se subclonaron en el vector pSANG10-3F tras obtener la secuencia de las scFv a partir de pET28b-scFv, para ello, se digirieron los vectores recombinantes con las enzimas de restricción *NcoI* y *NotI*, presentes en el sitio de clonaje múltiple del vector pSANG10-3F. Este vector contiene la secuencia pelB, que favorece la secreción de las scFv al periplasma.

La expresión de las scFv fusionadas al triple FLAG se llevó a cabo en *E. coli* BL21 (DE3). Tras inducir la expresión con IPTG se valoraron la capacidad de reconocimiento de las scFv presentes en el periplasma mediante ELISA frente a los antígenos purificados. En la Figura 29B se observa que en todos los casos, excepto la scFv F9 antiCXCL3, el reconocimiento del antígeno fue positivo aunque en ningún caso el título fue muy elevado.

En la figura 29C se muestra el análisis de expresión y solubilidad de las scFv. Tres scFv, E12 antiCXCL3, D4 antiCXCL1 y B2 antiMMP7, se sobreexpresan en la fracción soluble. Las 5 restantes fueron insolubles tras la introducción del 3xFLAG. La presencia de 11 residuos de ácido aspártico en el 3xFLAG puede favorecer la insolubilidad de la proteína de fusión. Las 3 scFv solubles se purificaron mediante cromatografía de afinidad utilizando una resina de Co^{2+} , obteniendo scFv con un nivel de pureza muy alto (Figura 29D).

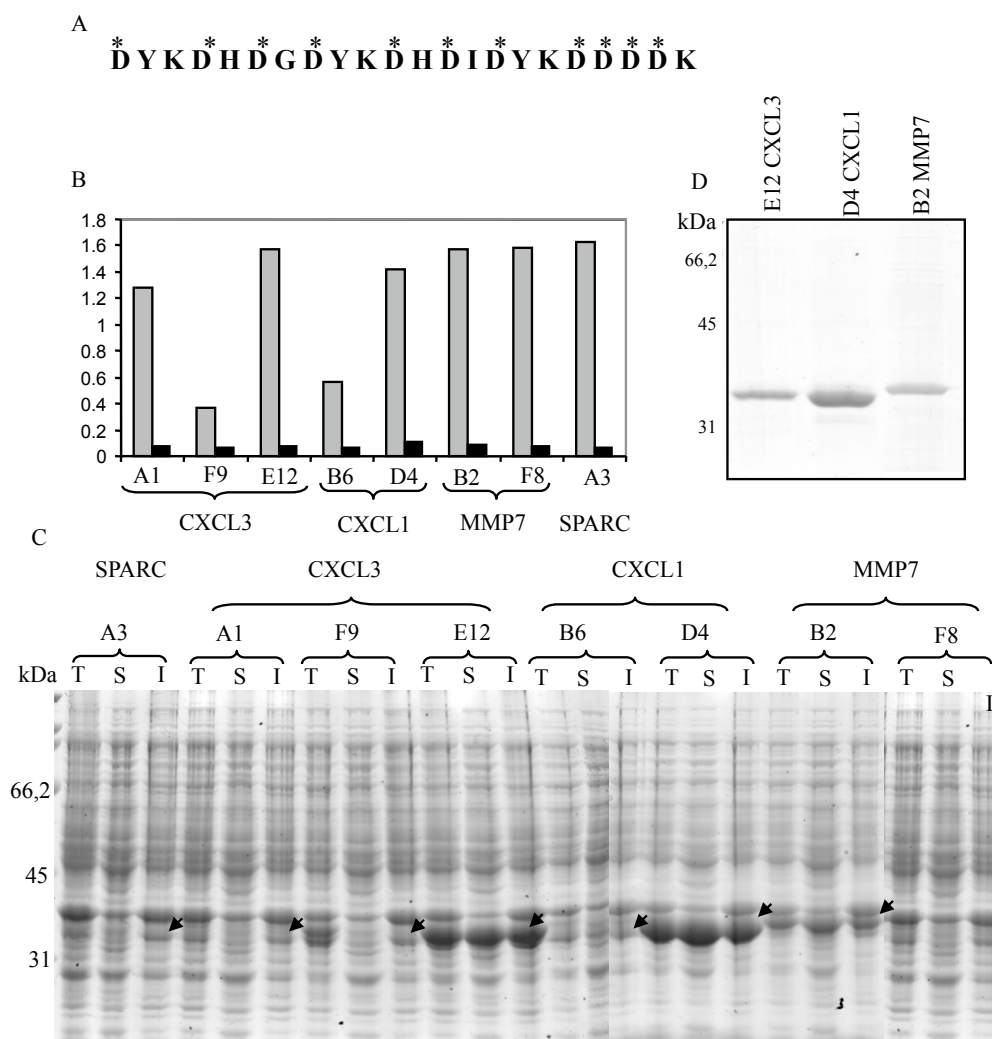


Figura 29. Expresión y solubilidad de las scFv seleccionadas a partir de la librería Mehta con 3xFLAG. A) Secuencia aminoacídica de 3xFLAG. El asterisco señala los residuos de aspártico de la secuencia. B) ELISA de periplasmas frente a los antígenos y *tags* purificados. C) Tinción con Azul brillante de *Coomassie* de las fracciones soluble e insoluble de cada clon. T: extracto total, S: fracción soluble, I: fracción insoluble. Las flechas señalan las bandas de sobre-expresión de las scFvs. D) Purificación de las scFv fusionadas a 3xFLAG con la resina Talon.

3.1.2.6 Expresión y purificación de las scFv fusionadas a 3xFLAG y tiorredoxina

Para comprobar si la insolubilidad ocasionada por el 3xFLAG se podría revertir, las scFv insolubles tras expresarlas en el vector pSANG10-3F, B6 antiCXCL1, A1 y F9 antiCXCL3, F8 antiMMP7 y A3 antiSPARC, se subclonaron en el vector pET32b, que contiene tiorredoxina en el extremo amino-terminal. Se ha descrito que la tiorredoxina actúa como agente solubilizante de las proteínas a las que se fusiona (Hammarström *et*

al, 2002). Además, pET32b proporciona un *tag* de histidinas en el extremo carboxilo-terminal de la proteína lo que favorece la posterior purificación de las scFv.

En la Figura 30A se muestra el análisis de expresión y solubilidad de las scFv portadoras de los *tags* 3xFLAG y Trx en la cepa de *E. coli* Origami (DE3) pLys. Todas las scFv fueron parcialmente solubles, confirmándose la capacidad solubilizadora de la tiorredoxina en bacteria. Sin embargo, un alto porcentaje continuó siendo insoluble.

A continuación, las scFv fueron purificadas por IMAC obteniendo en todos los casos unos niveles de pureza y concentraciones muy elevados (Figura 30B). En la Figura 30C se muestra el resultado de la digestión con enterocinasa de la scFv-3xFLAG fusionadas a Trx (tiorredoxina) F9 anti CXCL3. Tras la digestión, se purificaron las scFv mediante intercambio aniónico en columnas HiTrapQ. Las scFv precipitaron al eliminar la tiorredoxina, confirmando nuevamente que el *tag* 3xFLAG insolubiliza las scFv.

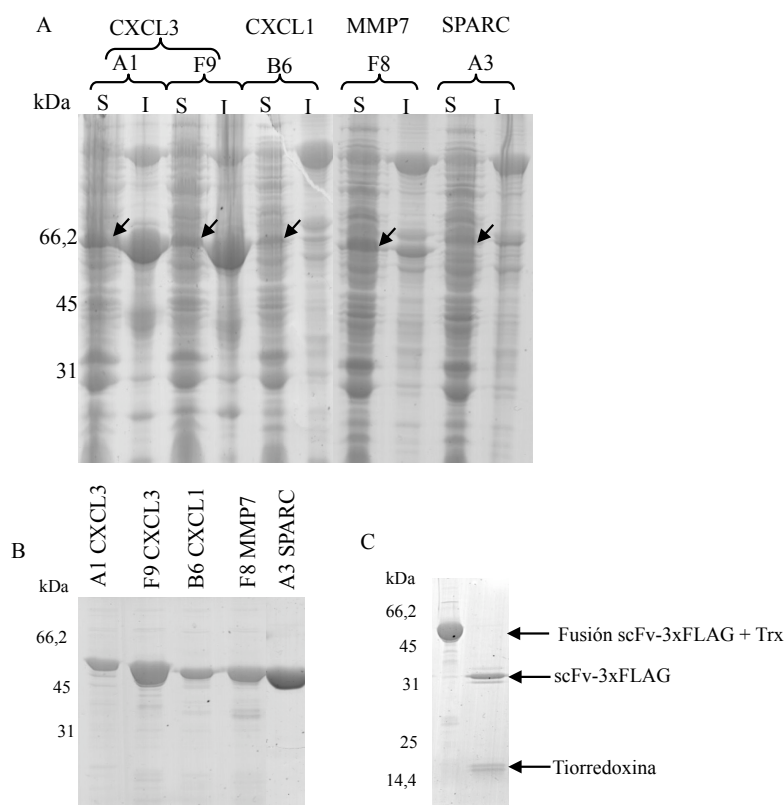


Figura 30. Purificación mediante IMAC y digestión de las scFv-triFLAG fusionadas a tiorredoxina. A) Tinción con Azul brillante de *Coomassie* de las fracciones de elución de la resina Talon. B) scFvs fusionadas a los *tags* 3xFLAG y Tiorredoxina purificadas mediante IMAC. C) Digestión de la scFv-3xFLAG-Trx antiCXCL3 con la enzima enterocinasa.

3.1.3 Estudio de la afinidad de las scFv seleccionadas mediante resonancia del plasmón de superficie

Para llevar a cabo la determinación de las afinidades de las scFv seleccionadas se utilizó la técnica de la resonancia del plasmón en superficie (SPR). Para el análisis de las constantes cinéticas de cada scFv se inyectaron en el Biacore X cinco concentraciones de cada scFv desde 1 μ M hasta 62,5 nM, disminuyendo la concentración en cada inyección a la mitad (Figura 31). Para el análisis se utilizó el modelo 1:1 Langmuir de interacción ente dos moléculas que supone una interacción equimolar entre las scFv y los antígenos. Se analizaron todas las scFv Mehta que se expresaron, es decir, fusionadas solamente a histidinas, con 3xFLAG y con tiorreoxina más 3xFLAG. En la Tabla 10 se muestran las constantes de asociación, disociación y de afinidad obtenidas.

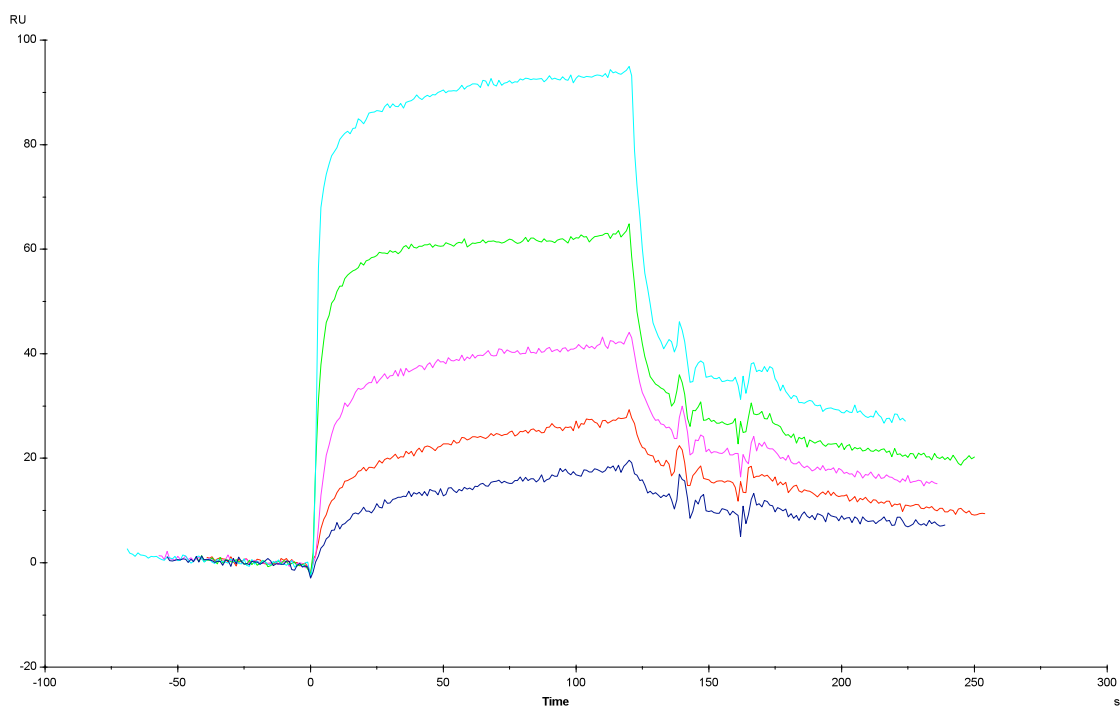


Figura 31. Sensorgrama de la interacción entre las scFv y los antígenos fijados en el chip. Se inyectaron 5 concentraciones diferentes de cada scFv para analizar las constantes cinéticas de cada una de ellas. En la figura se muestran los sensorgramas obtenidos con la scFv A1 anti CXCL3.

Tabla 10. Constantes cinéticas de las scFv seleccionadas a partir de las librerías Mehta.

scFv	Antígeno	Tag	kon (1/Ms)	koff (1/s)	K _D (μM)
B6	CXCL1	N	3,17E+04	5,00E-03	0,158
B6	CXCL1	3xFLAG-Trx	2,76E+03	4,86E-02	17,6
D4	CXCL1	N	2,14E+04	7,60E-03	0,355
D4	CXCL1	3xFLAG	x	x	
A1	CXCL3	N	1,62E+05	4,60E-03	0,0284
A1	CXCL3	3xFLAG-Trx	9,96E+04	4,90E-03	0,0492
E12	CXCL3	N	4,64E+04	5,58E-03	0,120
E12	CXCL3	3xFLAG	x	x	
F9	CXCL3	N	3,80E+05	6,10E-03	0,016
F9	CXCL3	3xFLAG-Trx	9,55E+02	1,98E-02	207
B2	MMP7	N	6,22E+03	1,27E-02	20,4
B2	MMP7	3xFLAG	x	x	
F8	MMP7	N	3,80E+03	2,95E-02	7,76
F8	MMP7	3xFLAG-Trx	3,59E+04	5,70E-03	0,159
A3	SPARC	N	5,80E+05	3,04E-03	0,0524
A3	SPARC	3xFLAG-Trx	5,76E+04	1,33E-02	0,231

N: scFv en vector original con histidinas

X: Constantes cinéticas que no pudieron ser determinadas

No se pudieron determinar las constantes cinéticas de las scFv fusionadas solamente al triple FLAG. La presencia del triple FLAG aporta una carga neta negativa a las scFv y la superficie de dextrano está cargada negativamente por lo que la ausencia de interacción entre los antígenos y las scFv puede deberse a una repulsión electrostática entre ellas.

Aunque las constantes de afinidad de las scFv Mehta originales fueron en general del orden nanomolar, especialmente A1, F9 y A3, la presencia de 3xFLAG y tiorredoxina modificó la capacidad de unión de las scFv a los antígenos disminuyendo, generalmente, la constante de afinidad en al menos un orden de magnitud con respecto a las scFv “originales” fusionadas tan sólo a histidinas. En el caso de la scFv F9 anti CXCL3 la disminución de la constante de afinidad al fusionar al anticuerpo los tags 3xFLAG y Trx es de hasta 4 órdenes de magnitud. Sorprendentemente, la scFv F8 anti MMP7 es la única que no sólo no disminuye su K_D al fusionarle los tags sino que mejora un orden de magnitud.

En resumen, la tecnología del despliegue en fagos nos permitió obtener anticuerpos recombinantes específicos frente a las 4 proteínas seleccionadas. Se

utilizaron dos librerías de fagos diferentes para llevar a cabo el proceso de selección de fagos, las librerías Tomlinson y las librerías Mehta. En nuestro estudio, las librerías Mehta resultó ser mucho más efectiva a la hora de obtener anticuerpos dado que la ausencia de codones *stop* TAG en la secuencia codificadora de las scFv permite expresar las scFv sin la proteína viral pIII fusionada al utilizar cepas no-supresoras del codon *stop* TAG, evitando la mutagénesis dirigida llevada a cabo con las scFv seleccionadas a partir de las librerías Tomlinson y los anticuerpos se obtienen en mayor cantidad y con una mejor afinidad.

3.2 Anticuerpos monoespecíficos basados en PrEST

En aquellos casos en que los niveles de expresión de las proteínas recombinantes fueron bajos o nulos y no se dispuso de las proteínas para la obtención de anticuerpos se expresaron fragmentos de las proteínas denominados PrEST correspondientes a secuencias únicas para la selección de anticuerpos. Todos ellos (excepto SPARC que se utilizó como control para llevar a cabo un estudio comparativo entre todos los diferentes métodos utilizados para la obtención de anticuerpos) presentaban un peso molecular superior a 50 kDa y habían sido inexpresables con el sistema *Gateway* descrito.

Se seleccionaron 5 proteínas para diseñar los PrEST y obtener anticuerpos monoespecíficos: AGT, EPHB3, PLAU, PTK7 y SPARC. Las proteínas diana se seleccionaron en base a su interés en la progresión tumoral y a la dificultad de expresarlas como proteínas completas en *E. coli*.

Los fragmentos se diseñaron de forma que i) contuviesen regiones hidrofílicas de la proteína para favorecer la solubilidad en *E. coli*, ii) fuesen lo suficientemente extensos como para presentar epítomos conformacionales y no sólo de secuencia o lineales y iii) se tratase de una secuencia única de entre todas las proteínas humanas conocidas para evitar reacciones cruzadas de los anticuerpos seleccionados frente a otras proteínas humanas (Agaton *et al*, 2003). En la Figura 32 se muestra el esquema general del proceso.

En el caso de las 5 proteínas seleccionadas se seleccionó un fragmento proteico que osciló entre 102 y 147 aminoácidos. En la Figura 33 se muestra el tamaño de los fragmentos proteicos expresados en relación al tamaño total de la proteína y su posición dentro de la secuencia.

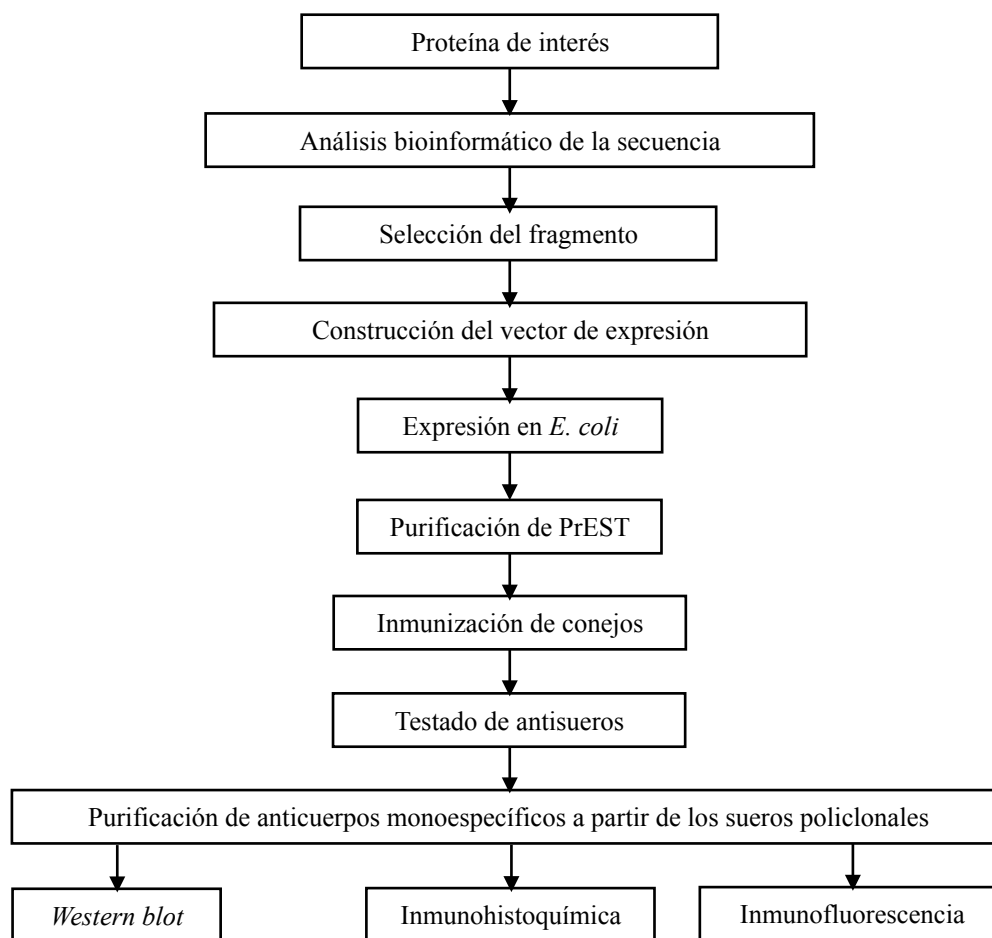


Figura 32. Esquema del proceso de obtención de anticuerpos monoespecíficos mediante la técnica PrEST.

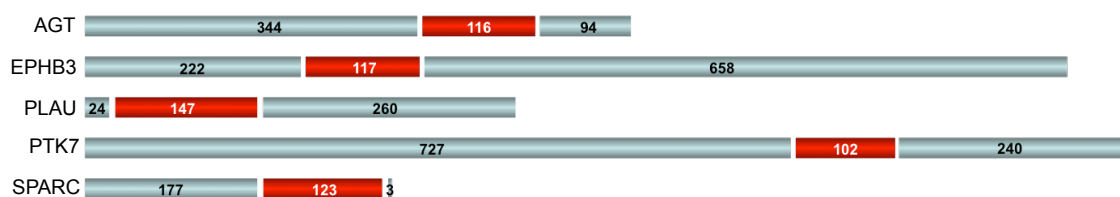


Figura 33. Fragmentos proteicos seleccionados para la obtención de anticuerpos monoespecíficos. Los fragmentos seleccionados de cada proteína para la inmunización de conejos y obtención de anticuerpos monoespecíficos se muestran en rojo en relación al tamaño total de la proteína diana. En cada fragmento se indica el número de aminoácidos que lo forman.

3.2.1 Expresión y purificación de los fragmentos proteicos PrEST en *E. coli*

Se diseñaron oligonucleótidos para amplificar las secuencias codificadoras de los 5 fragmentos seleccionados a partir de los clones de cDNA de secuencia completa de la MGC (AGT, EPHB3, PLAUI, PTK7 y SPARC). En los cinco casos analizados se logró construir satisfactoriamente el vector de expresión al introducir los productos de la PCR (llevada a cabo con la enzima de alta fidelidad Platimun Pfx DNA polimerasa) en

el vector pET15b, que presenta únicamente un *tag* de histidinas, tras digerir scFv y vector con las enzimas de restricción NdeI y BamHI.

La expresión de las proteínas se indujo con IPTG en células de *E. coli* BL21 (DE3) transformadas con los vectores contruidos. Para la lisis celular se utilizó el tampón de equilibrado fosfato sódico 20mM/ NaCl 0,5M/ urea 6M. En la Figura 34A se observa la sobreexpresión de cuatro de las cinco proteínas expresadas, AGT, EPHB3, PTK7 y SPARC. En el caso del fragmento de PLAU no tiene lugar una sobreexpresión aparente. Los fragmentos proteicos se purificaron con columnas de Ni²⁺ utilizando como tampón para el equilibrado de la columna el mismo utilizado para la lisis celular. Las concentraciones proteicas finales obtenidas tras purificar y dializar los PrEST hasta una concentración final de urea de 1M fueron de 2,7 mg/ml en el caso de AGT; 0,389 mg/ml en el caso de EPHB3; 1 mg/ml en el caso de PTK7 y 0,916 mg/ml en el caso de SPARC. Aparentemente no se detectan productos de degradación de los fragmentos expresados, obteniendo un grado de pureza muy elevado en todos los casos (Figura 34B).

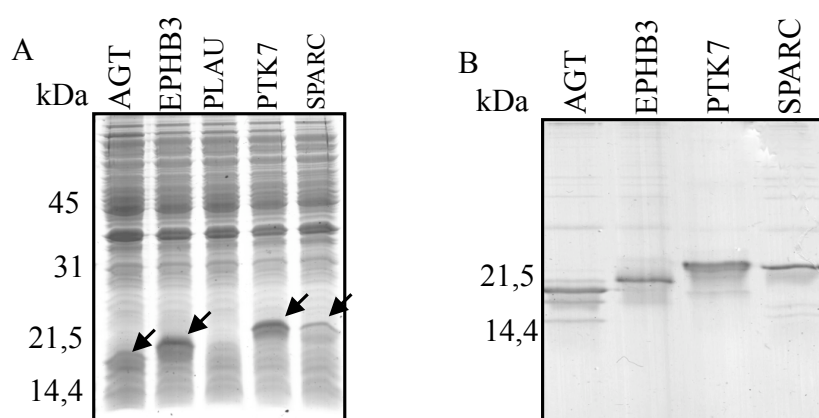


Figura 34. Expresión y purificación de los fragmentos proteicos seleccionados para la obtención de anticuerpos mono-específicos. Los fragmentos seleccionados se clonaron en el vector de expresión pET15b y se transformaron células BL21 (DE3). A) Análisis de la expresión de los fragmentos en bacteria mediante tinción con Azul brillante de *Coomassie* de tras lisar bacterias con tampón de equilibrado conteniendo urea 6M. Las flechas señalan las bandas correspondientes a cada PrEST. C) Fragmentos PrEST purificados.

En condiciones normales un alto porcentaje de PrEST es insoluble, además, el proceso de resolubilización de las proteínas expresadas lleva mucho tiempo si se trata de obtener los productos proteicos de forma rápida y a gran escala; por ello, todo el proceso de lisis celular y purificación de proteínas se hizo en condiciones

desnaturalizantes con urea 6M, diluyéndola luego hasta 1M para la inmunización de los conejos.

3.2.2 Obtención de anticuerpos monoespecíficos frente a PrEST

Para la obtención de anticuerpos frente a los PrEST se inmunizaron dos conejos con cada uno de dichos fragmentos. Los sueros obtenidos tras cada inyección de antígeno se titularon mediante ELISA (Figura 35). En cada caso se eligió el suero del conejo con mejor título para purificar los anticuerpos monoespecíficos. En el caso de EPHB3 no se consiguió obtener buenos sueros funcionales, posiblemente debido a que la concentración proteica fue menor y la cantidad neta de urea inyectada fue mayor que en resto de los casos.

Los sueros policlonales obtenidos de la inmunización de animales no suelen contener más de 1% de anticuerpos específicos frente al antígeno que ha sido inyectado (Uhlén & Pontén, 2005), por eso se decidió purificar los anticuerpos monoespecíficos a partir de los sueros policlonales.

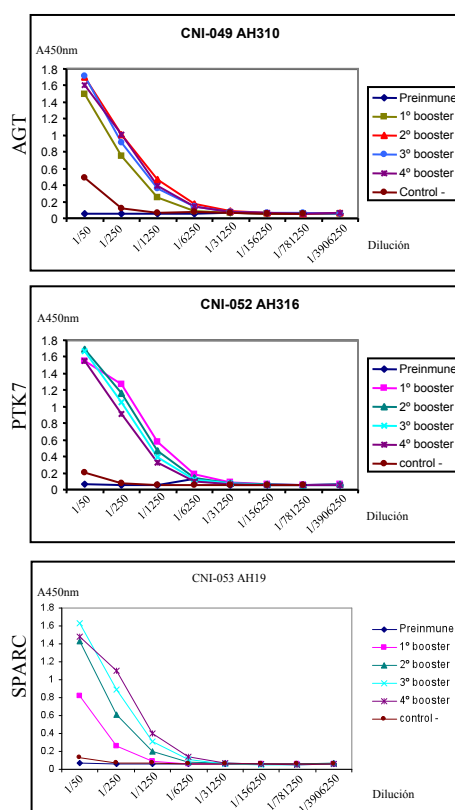


Figura 35. Titulación mediante ELISA de los sueros policlonales de los conejos inmunizados con los PrEST. Se utilizaron diluciones decrecientes de los sueros de los conejos inmunizados con los fragmentos proteicos para titular los anticuerpos policlonales generados frente a los PrEST. Se fijaron 0,3 µg de cada PrEST en las placas de ELISA. Como control negativo se hicieron reacciones cruzadas entre los diferentes sueros.

La purificación de los anticuerpos monoespecíficos a partir de los sueros se llevó a cabo mediante cromatografía de afinidad en columnas construidas específicamente para cada fragmento utilizando cada uno de ellos como ligandos de los anticuerpos monoespecíficos presentes en el suero policlonal (Agaton *et al*, 2004) (Figura 36A). Como se observa en la Figura 36B, los anticuerpos monoespecíficos purificados muestran un nivel de pureza razonable y unas concentraciones de 1,53 mg/ml en el caso de los anticuerpos monoespecíficos frente a AGT, de 1,13 mg/ml para los anticuerpos frente a PTK7 y de 1,64 mg/ml en el caso de los anticuerpos frente a SPARC.

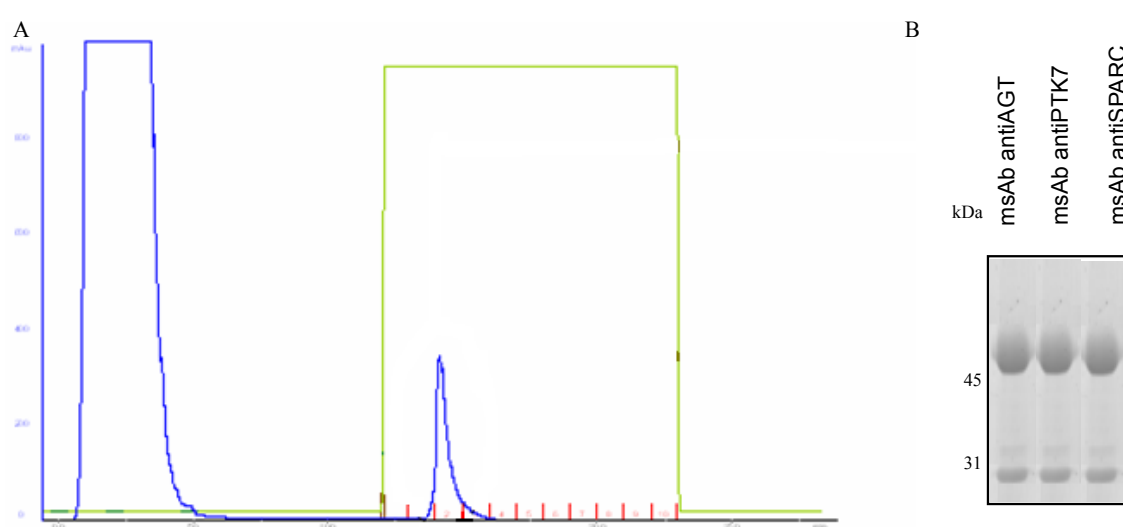


Figura 36. Purificación de los anticuerpos monoespecíficos mediante cromatografía de afinidad. Se construyeron columnas de afinidad específicas para cada suero con los fragmentos proteicos purificados. A) Cromatograma de la purificación de los anticuerpos monoespecíficos. Se muestra la purificación de los mAbs frente a AGT. La línea azul muestra la absorbancia a 595nm, la línea verde refleja el porcentaje de tampón de elución. B) Anticuerpos monoespecíficos purificados.

Los anticuerpos monoespecíficos purificados se titularon mediante ELISA. En la Figura 37 se observa que los anticuerpos purificados reconocen específicamente el fragmento frente al que fueron inmunizados los conejos. Los anticuerpos purificados se utilizaron posteriormente en estudios inmunohistoquímicos y *western blot*.

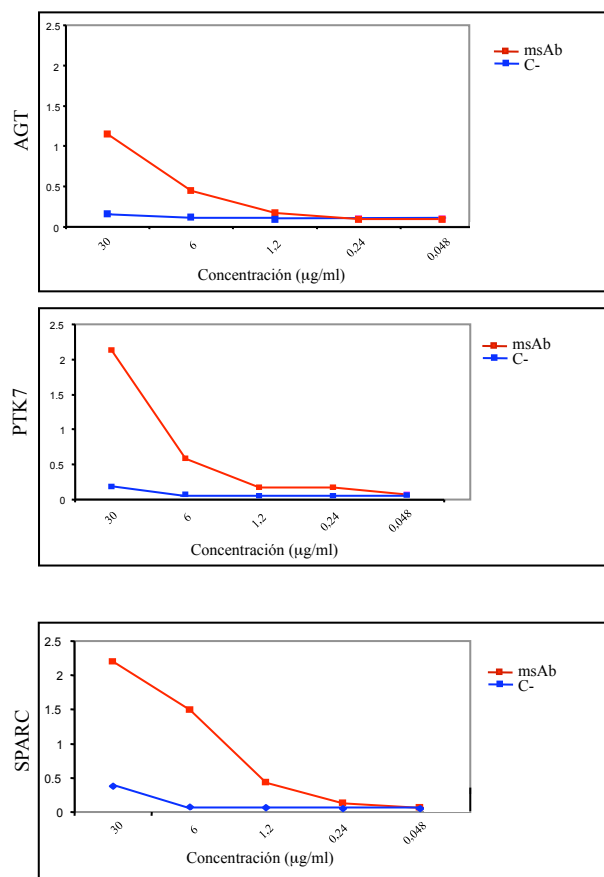


Figura 37. Titulación por ELISA de los anticuerpos monoespecíficos. Se utilizaron diluciones decrecientes de los anticuerpos monoespecíficos generados frente a los PrEST. Se recubrieron las placas de ELISA con 0,3 µg de cada PrEST. Como control negativo se hicieron incubaciones cruzadas entre los anticuerpos y los antígenos. Línea roja: anticuerpos monoespecíficos; línea azul: control negativo.

3.3 Anticuerpos monoclonales y policlonales

Las proteínas expresadas en bacteria también se utilizaron para la obtención de anticuerpos mediante las técnicas clásicas de producción de anticuerpos monoclonales y policlonales.

Para la producción de anticuerpos monoclonales se utilizaron las proteínas recombinantes de secuencia parcial BMP4 fusionada a MBP y LCN2 fusionada a GST para inmunizar ratones Balb/C. Una vez construidos los hibridomas y seleccionado el mejor clon, se utilizaron directamente los sobrenadantes de los cultivos celulares para llevar a cabo los experimentos de inmunocitoquímica (Figura 38), *western blot* e inmunohistoquímica. Los anticuerpos monoclonales obtenidos fueron funcionales en ambas técnicas.

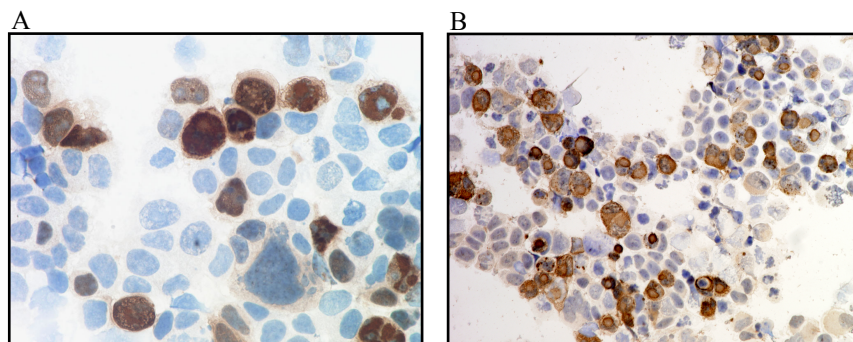


Figura 38. Inmunocitoquímica con los anticuerpos monoclonales antiBMP4 y LC2. Células Cos-1 transfectadas con A) BMP4-pcDNA3.2/nV5-DEST (aumento 20x) y B) LCN2-pcDNA3.2/nV5-DEST (aumento 10x) tras la tinción con el anticuerpo monoclonal específico.

Para la obtención de anticuerpos policlonales se utilizó la proteína recombinante de secuencia completa ANXA3 fusionada al *tag* de histidinas para inmunizar conejos *New Zealand White*. Los anticuerpos específicos frente a ANXA3 se purificaron a partir del suero de los conejos mediante cromatografía de afinidad en columnas de Proteína A.

4. Validación de las dianas seleccionadas en CCR

4.1 Análisis de los niveles de expresión proteica en cáncer colorrectal mediante *western blot*

Para confirmar que las proteínas seleccionadas se sobreexpresan en cáncer colorrectal se realizaron *western blot* tras resolver por electroforesis en geles de poliacrilamida-SDS los extractos proteicos de muestras normales y tumorales procedentes de seis pacientes con CCR en estadios Dukes de A a C y de diferentes líneas celulares de cáncer de colon (Colo205, Sw48, Sw480, Hct15 y Hct116).

En la Figura 39 se muestran los *western blot* realizados con las muestras tumorales y normales. Se observa que el anticuerpo policlonal frente a ANXA3 y los anticuerpos monoclonales frente a MMP7 y SPARC fueron especialmente eficientes en la diferenciación entre las muestras tumorales y normales, sin marcar preferencias entre los distintos estadios de la progresión tumoral. El anticuerpo monoclonal frente a BMP4 reconoce preferentemente los estadios Dukes B y C aunque las muestras normales también mostraban niveles de expresión significativos. En el caso de LCN2 se observa un incremento de la expresión a medida que la progresión tumoral avanza hacia los

estadios Dukes B y C. Los niveles de expresión de MMP11 permiten discriminar entre muestras tumorales y normales, especialmente en las muestras de estadio Dukes C. La única proteína que no pudo ser validada por *western blot* en muestras tumorales fue SPP1, que mostró un patrón de distribución anómalo en todos los pares de muestras. Como control positivo para la normalización de los datos se utilizaron los niveles de tubulina.

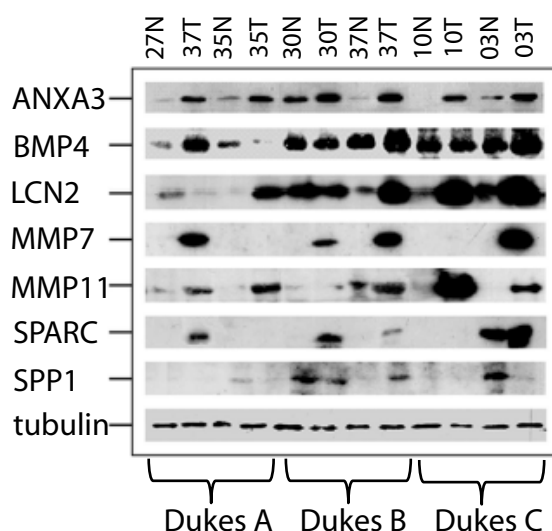


Figura 39: Análisis de la expresión de las dianas de cáncer colorrectal seleccionadas en muestras de tejido. El estudio se realizó tanto con anticuerpos que se desarrollaron anticuerpos (ANXA3, LCN2 y BMP4) como con anticuerpos disponibles comercialmente (SPARC, SPP1 y MMP11). Como control se utilizó antiTubulina. Se cargaron geles SDS-PAGE al 10% con extractos proteicos de tejidos normales (N) y tumorales (T) emparejados de 6 pacientes de cáncer colorrectal (dos de cada estadio Dukes A-C). Se transfirieron las proteínas a membranas de nitrocelulosa y se incubaron con los anticuerpos específicos. El revelado se hizo mediante quimioluminiscencia (ECL) o SuperSignal Femto (Pierce).

En el caso de las líneas celulares (Figura 40), los anticuerpos monoespecíficos frente a AGT dieron lugar a bandas de menor tamaño al esperado. Los *western blot* con los anticuerpos monoespecíficos frente a PTK7 mostraron, al igual que en el caso de AGT, bandas de menor tamaño al esperado. En el caso de SPARC, se observa expresión de esta proteína en las líneas Colo205, Hct15, Hct116, Sw48 y Sw480, siendo Sw480 la que muestras los mayores niveles de expresión. En la línea RKO no se observa expresión de SPARC

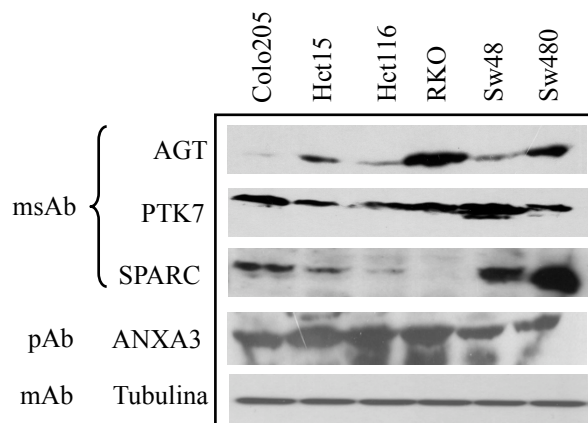


Figura 40. Western blot frente a las proteínas AGT, SPARC y PTK7 con los anticuerpos monoespecíficos. Se transfirieron a las membranas de nitrocelulosa los extractos proteicos totales de las líneas de cáncer de colon indicadas. Como anticuerpo primario se utilizaron los anticuerpos obtenidos frente a las proteínas indicadas. Pesos moleculares, en kDa, AGT: 58,9; PTK7: 118,3; SPARC: 44,6; ANXA3: 36,4; Tubulina: 49,9. msAb: anticuerpo monoespecíficos; pAb: anticuerpo policlonal; mAb: anticuerpo monoclonal.

4.2 Análisis de los niveles de expresión proteica en cáncer colorrectal mediante inmonohistoquímica

Se analizó la expresión de 11 proteínas seleccionadas (ANXA3, BMP4, CA9, IGFBP3, LCN2, MMP7, MMP11, PTK7, SPARC, SPP1 y TDGF1) en tejidos tumorales mediante inmunohistoquímica con un *microarray* de tejido específico de CCR conteniendo 97 muestras humanas emparejadas (normal-tumoral). Las imágenes obtenidas se muestran en la Figura 41.

Anexina A3 muestra una clara sobreexpresión de esta proteína en los tejidos tumorales observándose un aumento en la intensidad de la tinción en el 63% de los casos. En la mayoría de los casos aparece una reactividad específica de las células epiteliales que rodean las criptas y el *villi* intestinal (51%), mientras que en el resto de los tejidos la tinción se produce en las células del estroma. Aunque se detecta tinción citoplasmática de ANXA3, la tinción más clara tiene lugar en la membrana celular. En los tejidos normales se observa una leve expresión de esta proteína en células epiteliales.

BMP4 se detecta tanto en el citoplasma de las células tumorales como en la matriz extracelular que rodea las células de la mucosa intestinal, de acuerdo con su condición de proteína extracelular. En un porcentaje menor de células se observa

expresión en el núcleo, en concordancia con la translocación descrita de esta proteína al núcleo para la activación de la vía SMAD (Balemans et al, 2002). Los tejidos normales muestran tinción en ciertas áreas del estroma pero no en las células epiteliales.

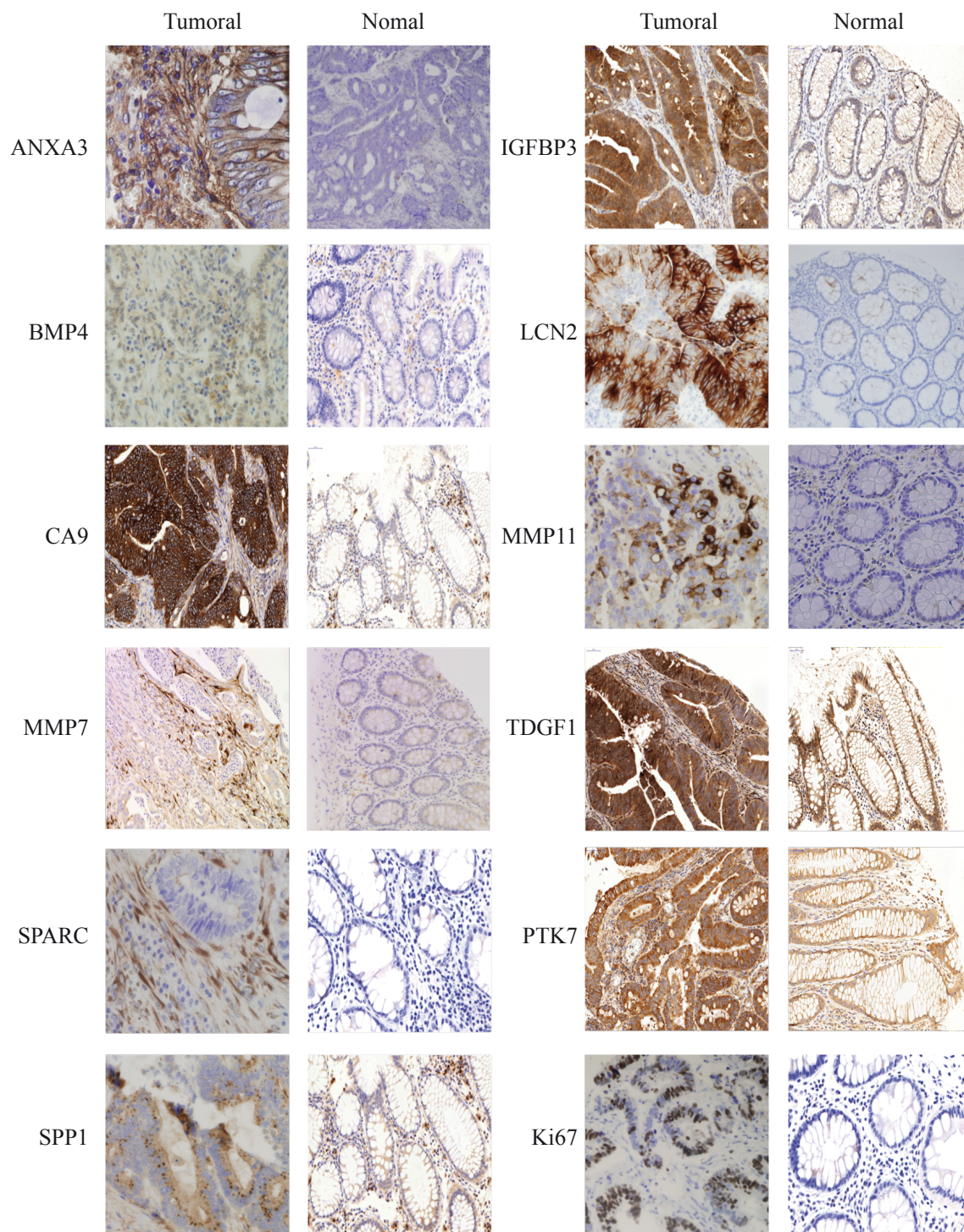


Figura 41. Análisis de la expresión de las dianas de cáncer colorrectal seleccionadas mediante array de tejido. En las imágenes se muestra el patrón de tinción de 11 proteínas seleccionadas (ANXA3, BMP4, CA9, IGFBP3, LCN2, MMP7, MMP11, PTK7, SPARC, SPP1 y TDGF1) en tejido tumoral y normal. Se utilizó como control Ki67, que tiñe células en proliferación, distintivo de tejidos neoplásicos.

CA9 se detecta mayoritariamente en el citoplasma y membrana plasmática de las células epiteliales pertenecientes al tejido tumoral, observándose una ligera tinción del estroma tumoral. No se observa expresión de esta proteína en las células epiteliales de la mucosa normal.

IGFBP3 presenta una distribución citoplasmática en las células epiteliales siendo los niveles de expresión en tejido tumoral mayores que en tejido normal.

El anticuerpo frente a LCN2 tiñe un alto número de células epiteliales en las muestras tumorales, la expresión es citoplasmática con refuerzo en la membrana. Las células positivas para LCN2 se distribuyen homogéneamente en el área luminal de los *microvilli*, aumentando la señal en el epitelio apical. No se observa expresión de esta proteína en la mucosa de los tejidos normales.

La expresión de MMP7 en las células tumorales muestra una distribución citoplasmática, con una fuerte señal en el aparato de Golgi, sobre todo en las zonas apicales de la neoplasia, en focos localizados al final de las criptas o en áreas de incipiente metástasis. Además de estos nódulos focales, MMP7 se expresa en el estroma que rodea las células epiteliales también en tejidos normales, dando al *microarray* una ligera tinción de fondo. La expresión de SPP1 aparece asociada a MMP7 mostrando un patrón de distribución similar en un pequeño porcentaje de los casos (13%), la expresión aparece en pequeños focos cerca de las cavidades de las glándulas o de los capilares sanguíneos. La señal en las células tumorales aparece en el citoplasma pero al tratarse de una proteína secretada se ve expresión en la matriz extracelular incluso en los tejidos normales.

MMP11 muestra una distribución peritumoral en las células del estroma de los tejidos neoplásicos, probablemente debido a su papel en la invasividad de las células tumorales. No hay expresión en los tejidos normales.

Los niveles de expresión de SPARC muestran un patrón similar al de MMP11 ya que se detecta una expresión muy clara en el estroma que rodea a las células tumorales en los tejidos neoplásicos, menor expresión en las displasias y ninguna en tejidos normales. SPARC se localiza en el citoplasma del estroma peritumoral.

La expresión de PTK7 aparece tanto en el núcleo como en el citoplasma de las células epiteliales así como en la matriz extracelular, presentando una débil tinción inespecífica en el tejido normal.

La tinción de TDGF1 en las células epiteliales del tejido tumoral es muy potente y muy superior a la observada en el tejido normal.

Como marcador positivo de proliferación se utilizó Ki67. Se observa una expresión muy intensa en las células neoplásicas de los tejidos tumorales, reflejo de su actividad proliferativa.

El análisis inmunohistoquímico confirmó una correlación positiva entre los niveles de transcripción y los niveles de expresión proteica en todos los casos, aunque algunos marcadores como SPP1, TDGF1 y PTK7 mostraron niveles de expresión significativos en tejidos de colon normal. Resultaron especialmente prometedores como marcadores tumorales ANXA3, CA9, LCN2, MMP7, MMP11 y SPARC.

5. Validación funcional de las dianas genómicas seleccionadas en CCR

Mediante la validación funcional de las dianas genómicas seleccionadas tratamos de establecer la implicación de las mismas en la apoptosis y proliferación celulares *in vitro*, utilizando para ello una serie de líneas celulares de cáncer de colon.

Por PCR cuantitativa se identificaron 3 genes que presentaban una expresión más elevada en las líneas celulares de cáncer de colon: ANXA3, BMP4 y MMP11. El efecto de estos genes en la apoptosis y/o proliferación tumorales y determinar si se trata de genes clave en la progresión de las células tumorales se estudió mediante RNA de interferencia en las líneas celulares de cáncer de colon Hct15 y Sw480 (Figura 14).

5.1 Estudio de la apoptosis tras la interferencia de los genes ANXA3, BMP4 y MMP11 en las líneas celulares de cáncer de colon Hct15 y Sw480

Para el silenciamiento de los genes diana se seleccionaron dos siRNA para cada gen. Los niveles de interferencia generados en las líneas celulares se analizaron por PCR cuantitativa a las 24, 48 y 72 h tras la adición de los siRNA para determinar cuál es el punto de mayor inhibición al que analizar los niveles de apoptosis inducidos por el silenciamiento de los genes. El mayor grado de interferencia de los tres genes se produjo 24 horas después de la interferencia en el caso de la línea Hct15 y a las 48h en el caso de Sw480. Estos tiempos se seleccionaron para el estudio de los niveles de apoptosis (Figura 42).

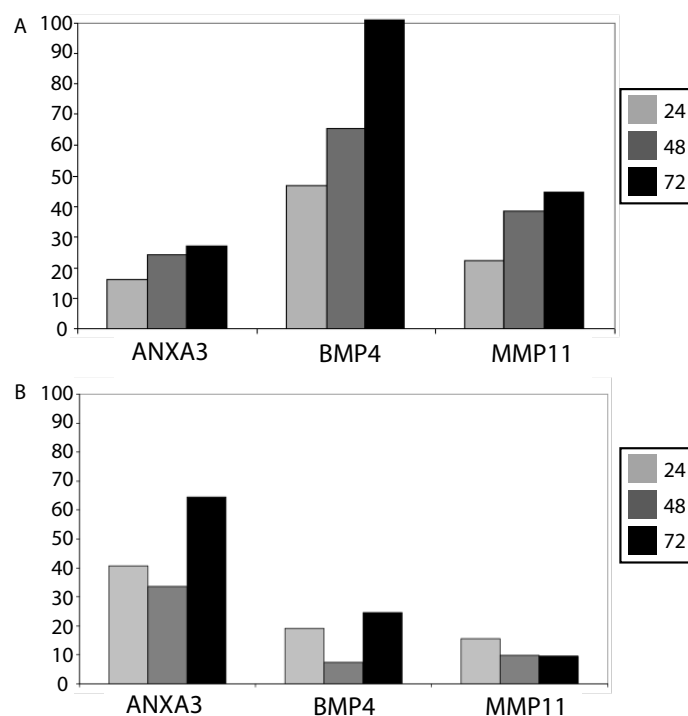


Figura 42. Niveles de interferencia de los genes ANXA3, BMP4 y MMP11 tras silenciar los mRNA con siRNA en las líneas celulares de cáncer de colon Hct15 y Sw480. Los niveles de interferencia se valoraron mediante PCR cuantitativa tras interferir con siRNA las líneas A) Hct15 y B) Sw480.

Para determinar los niveles de apoptosis inducidos por el silenciamiento de los genes diana se incubaron las células interferidas con AnexinaV, que se une a la fosfatidil serina que aparece expuesta en la membrana plasmática de las células apoptóticas, marcando la apoptosis temprana, e ioduro de propidio, que se une al DNA marcando la apoptosis tardía. Como se muestra en la Figura 43, en la línea Sw480 la interferencia de los genes ANXA3 y MMP11 inducen una moderada elevación en los niveles de apoptosis en la línea Sw480. En el caso de la interferencia de BMP4 se observa un aumento significativo en los niveles de apoptosis en relación al control. Las interferencias de la línea Hct15 no indujeron ningún cambio en los niveles de apoptosis de las células.

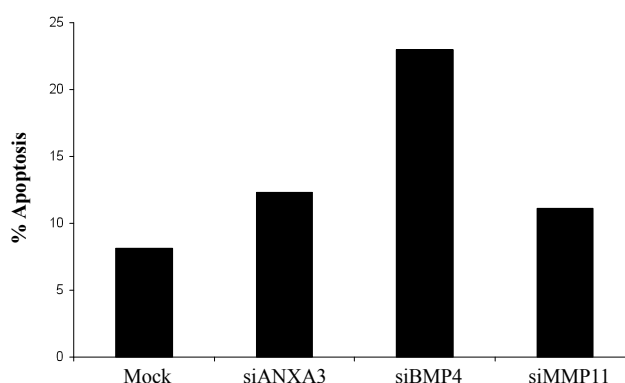


Figura 43. Análisis de los niveles de apoptosis tras silenciar con siRNA los genes ANXA3, BMP4 y MMP11 en la línea celular de cáncer de colon Sw480. Los niveles temprana se midieron con Annexina V. El porcentaje de apoptosis basal (valor 0) se estableció a partir del porcentaje de apoptosis en la población *wild type* tratada con oligofectamina.

5.2 Transfección de la línea celular Hct116 con los genes ANXA3 y BMP4 fusionados a GFP

La línea Hct116 no muestra sobreexpresión de RNA mensajero de los genes ANXA3 y BMP4 (Figura 14). Esta línea celular tiene una eficiencia de transfección por electroporación muy elevada (90-95%) por lo que fue elegida para el estudio de la influencia de estos genes en la proliferación celular al transfectar las células con dichos genes fusionados a la proteína GFP. Una vez transfectadas las células, se llevó a cabo un ensayo de proliferación celular durante 5 días consecutivos. Los resultados se muestran en la Figura 44 y reflejan que la introducción de ANXA3 no afecta a la proliferación celular. Sin embargo, la transfección de BMP4 sí ralentiza la proliferación de las células.

Tanto en los ensayos de apoptosis como en los de proliferación celular se observa que BMP4 ejerce un efecto sobre las células. En el caso de la línea celular Sw480, la interferencia de BMP4 promueve la apoptosis celular mientras que en Hct116 su sobreexpresión ralentiza la proliferación celular.

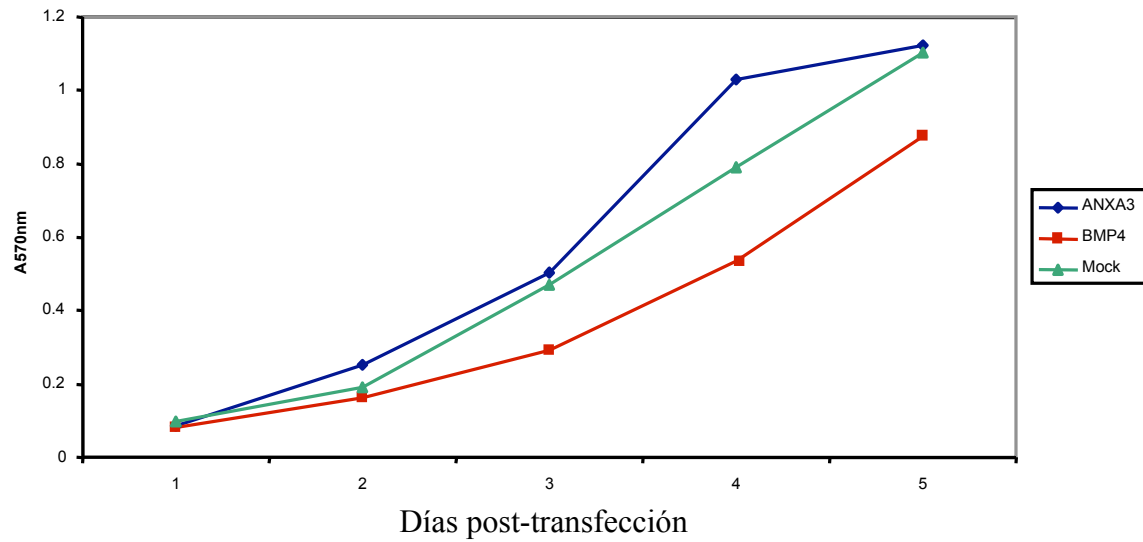


Figura 44. Ensayo de proliferación celular. La línea HCT116 se transfectó con los genes ANXA3 y BMP4 fusionados a GFP. Los niveles de proliferación celular se analizaron mediante un ensayo de MTT. La transfección de ANXA3 no afecta a la proliferación de las células mientras que BMP4 la ralentiza.

Discusión

El cáncer colorrectal (CCR) es el tipo de cáncer más frecuente detrás del cáncer de pulmón en hombres y de mama en mujeres por lo que el diagnóstico precoz y un tratamiento eficaz son dos de los mayores retos de la medicina en los países desarrollados. La rapidez en el diagnóstico y en el subsiguiente tratamiento depende en parte de la existencia de marcadores tumorales que permitan diferenciar un tejido sano de uno tumoral en las primeras etapas del desarrollo de la neoplasia. Nuestro laboratorio está interesado en la identificación de nuevos biomarcadores para mejorar el diagnóstico y pronóstico del cáncer colorrectal. Con este objetivo se llevó a cabo un análisis transcriptómico global para determinar la presencia de genes alterados en CCR.

Tras los análisis de expresión génica realizados con 22 pares de muestras de cáncer colorrectal se detectaron una serie de genes sobreexpresados en las muestras tumorales con respecto a las normales que podrían representar buenos marcadores de CCR. Numerosos estudios han puesto de manifiesto la gran cantidad de genes implicados en el desarrollo y la progresión tumorales pero muy pocos han tratado de averiguar si esas alteraciones en los niveles de expresión génica se ven reflejadas a nivel proteico y cuáles son los problemas a los que nos enfrentamos al realizar este proceso. En esta tesis, la estrategia utilizada para la validación de las dianas genómicas implicadas en la tumorigénesis colorrectal ha sido el uso de anticuerpos para determinar no sólo los niveles de expresión de las proteínas alteradas en la neoplasia sino también su localización celular y tisular. La selección de los genes para el estudio no estuvo sesgada por características como el peso molecular de sus productos proteicos o la función biológica de los mismos, centrándonos en proteínas de la membrana celular o de la matriz extracelular por ser más accesibles para el diagnóstico y terapia basados en anticuerpos.

Una de las limitaciones en la validación proteómica de los resultados obtenidos tras análisis genómicos es la disponibilidad de anticuerpos específicos frente a las proteínas diana para analizar la expresión proteica diferencial entre tejidos sanos y tumorales utilizando técnicas como la inmunohistoquímica o los *microarrays* de proteínas. Por ello, este trabajo de tesis doctoral parte de la producción de proteínas con potencial interés diagnóstico y terapéutico para utilizarlas como antígenos en la obtención de anticuerpos.

El primer paso en la producción de anticuerpos específicos frente a los genes seleccionados fue la obtención de las proteínas que se utilizaron como antígenos. Como sistema de expresión de las proteínas se optó por *Escherichia coli* por ser el modelo más rápido y más adecuado en los trabajos a gran escala. En primer lugar, fue necesario disponer de los cDNA a partir de los cuales construir los vectores de expresión de proteínas. Como fuente de cDNA se intentó utilizar los clones IMAGE utilizados para la impresión de los chips de cDNA pero, dado que muchos no alcanzaban el 30 % de la fase de lectura abierta del gen, se utilizaron como alternativa los clones de secuencia completa de la colección MGC. La obtención de los cDNA correspondientes a los 29 genes fue uno de los pasos más lentos de todo el proceso. Sin duda, la disponibilidad de una colección pública de clones de cDNA con la fase de lectura del gen completa y verificada en vectores compatibles con un sistema de recombinación como *Gateway* sería muy beneficioso para los proyectos de expresión de proteínas a gran escala (Brizuela *et al*, 2002).

Para la construcción de los vectores de expresión tanto a partir de la colección IMAGE como de la colección MGC se utilizó el sistema de vectores TOPO-*Gateway*. El sistema TOPO-*Gateway* utiliza la enzima topoisomerasa, que se une a los extremos libres del vector, y presenta la ventaja de que los oligonucleótidos necesarios para la recombinación son de un tamaño más pequeño; además, la eficiencia de transformación de células competentes con los vectores TOPO es muy elevada. En nuestro estudio, el 90% de las construcciones en pENTR se consiguieron al primer intento. Sin embargo, en el 10% de los clones restantes se observaron pequeñas deleciones que nos obligaron a secuenciar todas las construcciones para comprobar el estado de todos los vectores construidos. Por ejemplo, el vector pENTR de IFITM1 contenía una deleción interna de 5 nucleótidos. Los clones en los que se detectaron mutaciones o deleciones fueron descartados para las siguientes fases del proceso.

Otra limitación del sistema *Gateway* es la necesidad de realizar tres pasos de transformación en *E. coli*, uno para el vector de entrada (pENTR) y dos más para el vector de destino, uno en la cepa de *E. coli* DH5 α , para obtener un número de copias del vector elevado, y otro en la cepa BL21 (DE3), para llevar a cabo la expresión de las proteínas. Para acelerar el proceso de obtención de las cepas de expresión de proteínas (BL21), tratamos de eliminar el paso de transformación en DH5

transformando directamente BL21 con el producto generado por la reacción de la LR clonasa, obteniendo un 95% de clones positivos en BL21 al primer intento; el 5% restante se llevó a cabo de forma clásica, transformando primero DH5a.

Respecto a la expresión de proteínas se observó que la solubilidad de las proteínas disminuyó significativamente a medida que aumenta el tamaño de la proteína. En este estudio, solamente las proteínas con un peso molecular por debajo de 50 kDa se expresaron como proteínas solubles con al menos uno de los tres *tags* utilizados, lo que supone una clara limitación en este tipo de proyectos en los que se trata de expresar a gran escala un gran número de marcadores genómicos. Se ha intentado predecir el éxito en la expresión de una proteína a partir de las características de la familia a la que pertenece o de los dominios que presenta sin ningún resultado (Klein y Dhurjati, 1995).

A la vista de los resultados obtenidos, MBP es el *tag* más recomendable si se pretende obtener proteínas recombinantes solubles. Sin embargo, el hecho de que las proteínas fusionadas a MBP sean solubles no significa que el plegamiento de las mismas sea correcto o que sean funcionalmente activas (Sachdev *et al*, 1998; Martínez-Torrecuadrada *et al*, 2005). Por otro lado, se trata de un *tag* de un peso molecular elevado (42 kDa) por lo que los animales inmunizados con proteínas recombinantes fusionadas al *tag* MBP suelen generar una gran cantidad de anticuerpos específicos frente al *tag*, bajando considerablemente el título frente al antígeno seleccionado en el caso de los anticuerpos policlonales y aumentando la dificultad de generar un hibridoma específico frente a la proteína de interés en el caso de los anticuerpos monoclonales. Sin embargo, las proteínas solubles son mucho más ventajosas para la producción de anticuerpos dado que mantienen los epítomos conformacionales naturales, como se observa en los resultados obtenidos con los anticuerpos monoclonales anti BMP4 y LCN2, y no producen reacciones adversas en los animales de experimentación.

En general, GST o 6xHis no favorecen la solubilidad de las proteínas tanto como MBP. En estudios previos, utilizando la misma estrategia para obtener 27 proteínas de bajo peso molecular (por debajo de 20 kDa), los mejores porcentajes de solubilidad se obtuvieron cuando las proteínas se fusionaron a MBP o tiorredoxina (Hammarström *et al*, 2002). Sin embargo, la tiorredoxina no se puede purificar por

cromatografía de afinidad por lo que es necesario fusionar otro *tag* a la proteína de interés convirtiendo de nuevo a la MBP como la mejor opción a la hora de expresar proteínas solubles.

Para la obtención de los anticuerpos frente a las proteínas seleccionadas se utilizaron cuatro estrategias diferentes, por un lado, se generaron anticuerpos monoclonales mediante la tecnología del despliegue en fagos (en formato de scFv) y mediante la técnica clásica de los hibridomas. Por otro lado, obtuvimos anticuerpos policlonales de conejo y monoespecíficos, derivados de purificar los anticuerpos específicos frente al antígeno a partir de los sueros.

La tecnología del despliegue en fagos se ha convertido en un método muy versátil a la hora de obtener anticuerpos humanos tanto para investigación como para aplicaciones clínicas y de diagnóstico (Leath *et al*, 2004). Las librerías de anticuerpos recombinantes se han convertido en importantes herramientas para la obtención de anticuerpos a gran escala de forma rápida y simultánea.

En este trabajo se utilizaron dos librerías de fagos para la selección de scFv específicas. En el caso de las librerías Tomlinson todas las scFv seleccionadas presentaron codones *stop* en la secuencia codificadora del anticuerpo, que impide la expresión de las scFv como productos solubles. Estudios previos realizados con las librerías Tomlinson muestran la alta frecuencia de codones *stop* en las CDRs de las scFv seleccionadas (Hirose *et al*, 1998; Yan *et al*, 2004; Liu *et al*, 2005; Martínez-Torrecuadrada *et al*, 2005). Aunque la proporción habitual de codones *stop* en las scFv seleccionadas oscila entre un 10 y un 30%, en algunos casos esta proporción puede ascender al 70-100%. Esto se debe a que durante la elaboración de librerías sintéticas se llevan a cabo reordenamientos de segmentos V_H y V_L *in vitro* y PCR con oligonucleótidos aleatorios para introducir CDR artificiales y generar librerías con el mayor repertorio de anticuerpos posible. Para solucionar este problema y obtener scFv solubles se optó por expresar las scFv en la cepa de *E. coli* supresora TG1 tras mutar el codon *stop* TAG (reconocido como glutamina por la cepas supresoras) por el codon TAA, que sí lo reconocen como *stop* (Barderas *et al*, 2006). A pesar del éxito en la expresión y purificación de las scFv, los niveles obtenidos no fueron suficientes para trabajar con las scFv en ensayos posteriores. Las librerías Mehta resultaron más útiles a la hora de obtener anticuerpos específicos frente a las proteínas diana dado

que las scFv obtenidas no presentaron codones *stop* en su secuencia, y presentaron mayor afinidad al provenir de una librería de mayor tamaño.

En ambas librerías observamos que la funcionalidad de las scFv solubles era mucho menor que cuando forman parte de la partícula viral. Las scFv presentan dos puentes disulfuro intramoleculares, uno en la V_H y otro en la V_L, para adquirir el plegamiento correcto y mantener la capacidad de reconocimiento del antígeno (Glockshuber *et al*, 1992); la formación de dichos puentes disulfuro es el factor limitante a la hora de expresar scFv u otros anticuerpos recombinantes funcionales (solubles, correctamente plegados y capaces de reconocer el antígeno) en *E. coli* (Fernández, 2004). Se ha propuesto que la principal causa de la inactivación irreversible de las scFv es una estabilidad insuficiente de las mismas debida a la carencia de los puentes disulfuro intramoleculares (Reiter *et al*, 1994).

Las scFv seleccionadas a partir de las librerías Mehta, aunque presentaban una buena afinidad, se subclonaron en diferentes vectores para fusionarles diferentes *tags* que facilitan su detección. Por un lado, se utilizó una cola de histidinas para poder purificarlas mediante cromatografía de afinidad. Sin embargo, la cola de histidinas no es lo más apropiado para los ensayos inmunológicos por el grado de inespecificidad de sus anticuerpos. Por ello, se utilizó el 3xFLAG ya que se ha descrito que este *tag* es apropiado para la detección de scFv en inmunohistoquímica (Martin *et al*, 2006). Sin embargo, en nuestro estudio la secuencia del 3xFLAG indujo la insolubilidad de la mayoría de las scFv por acidificar considerablemente la proteína al añadir 11 residuos de aspártico. Para solubilizar las scFv fusionadas al triple FLAG se probó a fusionarlas con tiorredoxina, descrito como agente solubilizante de proteínas en *E. coli* (Hammarström *et al*, 2002). Para la obtención de las scFv solubles fusionadas a los *tags* 3xFLAG y Trx fue necesario utilizar la cepa de *E. coli* Origami (DE3) pLys, que facilita la formación de puentes disulfuro ya que presenta mutaciones en las enzimas tiorredoxina reductasa y glutation reductasa.

Con la combinación de los *tags* 6xHis, 3xFLAG y Trx se consiguió obtener de forma soluble las scFv fusionadas al triple FLAG y purificarlas mediante cromatografía de afinidad por unión del *tag* 6xHis. Sin embargo, las constantes cinéticas de las scfv seleccionadas en sus diferentes formatos disminuyeron en muchos casos respecto a las originales.

En nuestro estudio, las scFv seleccionadas resultaron altamente específicas al utilizarlas en *western blot*, ELISA y Biacore frente a los antígenos purificados, sin embargo, los resultados obtenidos en inmunohistoquímica fueron insatisfactorios, probablemente debido a su baja afinidad. Normalmente, los anticuerpos obtenidos a partir de librerías *naïve* o sintéticas presentan constantes de afinidad en el rango micromolar (Yau *et al*, 2005)), que no son suficientes para utilizar dichos anticuerpos como agentes de diagnóstico inmunohistoquímico o terapéuticos, donde es necesario alcanzar altas afinidades en el rango subnanomolar (Lowe y Jermutus, 2004). En comparación, los anticuerpos generados *in vivo* mediante la inmunización de animales pueden alcanzar afinidades de hasta 0.01 nM (Foote y Eisen, 1995).

La afinidad de las scFv se puede mejorar utilizando técnicas de evolución *in vitro* como la obtención de nuevas librerías a partir de un anticuerpo dado mediante la mutagénesis de sus CDR seguida de nuevas rondas de selección. La estructura del lugar de reconocimiento del antígeno determina tanto la especificidad del anticuerpo como su afinidad. Está formado por tres *loops* hipervariables (CDR) de la cadena pesada y tres de la cadena ligera, siendo el CDR3H el que mayor influencia tiene en el reconocimiento (Xu *et al*, 2000) por presentar el mayor grado de variabilidad. Por este motivo la CDR3H es el principal objetivo de la mutagénesis dirigida. Las mutaciones realizadas en la secuencia de soporte (*framework*) afectan principalmente a la estabilidad de la molécula, no a la afinidad de la misma.

En resumen, la selección de scFv mediante la tecnología del despliegue en fagos resultó ser un proceso lento (por los múltiples subclonajes necesarios para otorgarles las características adecuadas) y de bajo rendimiento para su aplicación en diagnóstico. La maduración de las scFv para aumentar su afinidad es un proceso largo y costoso si, como en nuestro caso, lo que se pretende es disponer de anticuerpos frente a un gran número de proteínas. Para que esta técnica supusiese una alternativa a la obtención de anticuerpos mono y policlonales para diagnóstico, sería necesario disponer de librerías de gran tamaño y de una automatización completa del proceso. Sin embargo, esta técnica sí puede ser útil cuando el objetivo es obtener anticuerpos terapéuticos frente a una proteína determinada.

Una de las maneras de acelerar el proceso de producción de anticuerpos es acelerar la obtención de las proteínas que actuarán como antígenos. El sistema de expresión de proteínas más rápido es *E. coli* pero, como se ha señalado anteriormente, presenta la desventaja, en general, de no producir proteínas heterólogas por encima de 50kDa (Madoz-Gúrpide y López-Serra *et al*, 2006). Para superar la dificultad de expresar proteínas de gran tamaño en *E. coli* se desarrollaron los PrEST (Agaton *et al*, 2003). Los PrEST son suficientemente grandes para, a diferencia de la inmunización con péptidos, obtener anticuerpos específicos frente a epítomos lineales y conformacionales tras la inmunización de animales. Los PrEST se están utilizando para realizar un Atlas del Proteoma Humano tanto de tejidos sanos como tumorales mediante inmunohistoquímica (Uhlén y Pontén, 2005). Para ello es necesario disponer de una batería de anticuerpos frente a todas las proteínas humanas e hibridarlos en cortes histológicos para analizar la distribución y abundancia de las proteínas en los diferentes tejidos humanos. La utilización de los anticuerpos monoespecíficos trata de evitar las limitaciones propias de los anticuerpos policlonales. Los anticuerpos policlonales reconocen múltiples epítomos del mismo antígeno pero presentan problemas de reactividad cruzada entre diferentes proteínas dando lugar a cierta inespecificidad de fondo en inmunohistoquímica. Por ello, la purificación de los anticuerpos específicos frente a un determinado antígeno a partir de un suero policlonal mediante cromatografía de afinidad utilizando el propio antígeno como agente de purificación permite evitar los problemas derivados de los anticuerpos policlonales.

Los anticuerpos monoespecíficos preparados frente a PTK7 y SPARC no fueron muy satisfactorios a la hora de determinar los niveles y patrón de expresión de las proteínas diana en tejidos sanos y tumorales en *microarrays* de tejido. A la vista de los resultados publicados en el Atlas del Proteoma Humano, (www.proteinatlas.org), parece que las limitaciones encontradas en nuestro trabajo son habituales en este tipo de estudios dado que las imágenes del Atlas muestran una alta variabilidad en el patrón de expresión de una determinada proteína en un determinado tejido para todas las proteínas analizadas hasta la fecha. Quizás sea necesario preparar varios PrEST de cada proteína hasta encontrar el más adecuado para la obtención de anticuerpos.

A la vista de los resultados obtenidos deducimos que los anticuerpos monoclonales derivados de hibridomas continúan siendo la mejor elección para conseguir resultados rápidos en diagnóstico. En los casos en lo que se utilizaron estos anticuerpos las imágenes de inmunohistoquímica obtenidas fueron mucho más limpias y específicas que con el resto de anticuerpos (scFv, policlonales y monoespecíficos). La mayor reactividad de los anticuerpos monoclonales en inmunohistoquímica se debe, entre otros motivos, a su mayor constante de afinidad y a la existencia de buenos anticuerpos secundarios.

Se seleccionaron las proteínas CXCL1, CXCL3, MMP7 y SPARC para la obtención y análisis por anticuerpos recombinantes humanos, AGT, EPHB3, PLAU, PTK7 y SPARC para la obtención y análisis por anticuerpos monoespecíficos y ANXA3, BMP4, CA9, IGFBP3, LCN2, MMP7, MMP11, SPARC, SPP1 y TDGF1 para la obtención y análisis por anticuerpos mono y policlonales convencionales. Las proteínas seleccionadas se pueden clasificar en grandes grupos teniendo en cuenta su funcionalidad y relación con la progresión tumoral.

Las proteínas CXCL1, CXCL2, LCN2 y SPP1 están relacionadas con los procesos inflamatorios que tienen lugar en los tejidos tumorales como consecuencia de la alteración de células mediadoras del sistema inmune (monocitos, neutrófilos, etc.). CXCL1 y CXCL3 pertenecen a la familia de las citoquinas, moléculas implicadas en el desarrollo, homeostasis y correcto funcionamiento del sistema inmune. Esta familia de proteínas también se ha relacionado con la angiogénesis, proceso fundamental en la progresión y metástasis de los tumores (Strieter *et al*, 1995). La Lipocalina 2 (LCN2) es una proteína de 25 kDa que se almacena dentro de gránulos en los neutrófilos (Cowland *et al*, 1997) y participa en la modulación de la inflamación. Se ha observado mediante inmunohistoquímica que la proteína LCN2 aparece sobreexpresada en tumores de mama, pulmón, colon y páncreas (Friedl *et al*, 1999). SPP1 (*Secreted Phosphoprotein 1*) es una de las citoquinas de respuesta inmune de tipo 1 mediada por macrófagos en ratón (Ashkar *et al*, 2000). SPP1 se ha identificado como uno de los genes diana del gen supresor de tumores P53 dado que sus niveles de expresión se ven aumentados por la actividad de p53 inducida por un daño en la secuencia de DNA, de hecho, el gen SPP1 presenta en su promotor un elemento de respuesta a p53, confirmando que se trata de una diana transcripcional de

la proteína p53 (Morimoto *et al*, 2002). SPP1 se ha descrito en diferentes tipos de cáncer como colorrectal (Irby *et al*, 2004), cáncer de ovario (Kim *et al*, 2002), cáncer de mama (Chakraborty *et al*, 2008) o cáncer gástrico (Tang *et al*, 2008).

Una serie de proteínas, MMP7, MMP11, SPARC, AGT, BMP4, PLAUI y CA9, están relacionadas con la invasión y metástasis de las células tumorales. MMP7 y MMP11 pertenecen a la familia de las metaloproteinasas de la matriz extracelular, enzimas implicadas en procesos de remodelación de los tejidos tanto en condiciones fisiológicas normales, desarrollo embrionario o remodelación de tejidos, como patológicas, como artritis y metástasis (Masaki *et al*, 2001), favoreciendo la expansión de las células en el tejido en el que se encuentran. MMP7 se sobre-expresa en diferentes tipos de cáncer humano y se relaciona con la metástasis al hígado a partir de un tumor colorrectal (Zeng *et al*, 2002). MMP11 se relaciona con el desarrollo de cáncer colorrectal en mujeres y la inestabilidad de microsatélites (Skoglund *et al*, 2002). SPARC (*Secreted protein, acidic and rich in cysteine*) es una glicoproteína multifuncional asociada a la matriz extracelular que modula la morfología de las células, inhibe la progresión del ciclo celular y regula la síntesis de la matriz extracelular (Bradshaw *et al*, 2003), su remodelación (Chlenski *et al*, 2006) y la permeabilidad del endotelio (Goldblum *et al*, 1994). Se ha descrito que esta proteína aparece sobreexpresada en los carcinomas de mama que han metastatizado al pulmón (Minn *et al*, 2005). AGT o Angiotensinógeno pertenece a la familia de los inhibidores de las proteasas de serpina. Se ha descrito que esta proteína está involucrada en la regulación de la hipertrofia de los miocitos (Shadosima *et al*, 1993) y que tanto AGT como sus derivados inhiben la angiogénesis (Célérrier *et al*, 2003). BMP4 es miembro de la familia de proteínas *Bone morphogenetic*, miembro de la superfamilia *Transforming growth factor-β*. Esta superfamilia incluye familias proteicas de factores de crecimiento y diferenciación. Las BMP utilizan receptores de BMP y proteínas SMAD intracelulares para transducir sus señales de diferenciación celular, proliferación y apoptosis. El alelo oncogénico de la β-catenina es necesario para la sobreexpresión y secreción de BMP4 por parte de líneas celulares de cáncer de colon humano con el gen *adenomatous polyposis coli* (APC) mutado, reflejando la relación entre la vía de señalización Wnt y la proteína BMP4 en el desarrollo tumoral (Kim *et al*, 2002). Por otro lado, se ha descrito un aumento en el grado de invasividad

de líneas celulares de cáncer colorrectal (Deng *et al*, 2007), de melanoma (Rothhammer *et al*, 2005) y gástrico (Hamada *et al*, 2007) cuando sobreexpresan la proteína BMP4, observándose en la morfología y grado de invasividad celulares. PLAU, o urokinasa, estimula la migración celular en células musculares humanas (Kiian *et al*, 2003) y se ha visto sobreexpresada en los focos invasivos del carcinoma de mama ductal (Nielsen *et al*, 2006). La Anhidrasa Carbónica IX (CA9) es miembro de la familia de metaloenzimas del zinc. Durante el desarrollo tumoral, las células se ven sometidas a diferentes tipo de estrés como hipoxia, falta de nutrientes o acidosis (debida a la acumulación de productos ácidos derivados del metaolismo glicolítico de las células). La adaptación a estas condiciones da lugar a la activación de una serie de genes implicados en la dispersión del tumor, entre ellos CA9, que participa en la regulación del pH celular, factor importante en condiciones de hipoxia y acidosis (Pastorekova *et al*, 2008). Se ha observado sobreexpresión de esta proteína en diferentes tipos de cáncer como colorrectal (Cleven *et al*, 2007), cerebral (Järvelä *et al*, 2008) o de mama y ovario (Hunakova *et al*, 2007).

Las proteínas ANXA3, EPHB3, PTK7, IGFBP3 y TDGF1 juegan un importante papel en la transducción de señales. Annexina A3 pertenece a la familia de las annexinas. Los miembros de esta familia proteica, proteínas de unión a fosfolípidos dependiente de calcio, juegan un papel importante en la regulación del crecimiento celular y en la transducción de señales. No hay ningún trabajo previo que relacione esta proteína con el proceso tumoral. EPHB3 (*Ephrin receptor EphB3*) pertenece a la familia de los receptores protein-tirosina kinasa. En el epitelio intestinal, la interacción de los receptores EPH con sus ligandos, las ephrinas, regula la proliferación y diferenciación celulares (Batlle *et al*, 2002). En los primeros estadios de la neoplasia colorrectal (criptas displásicas aberrantes y adenomas pequeños) los niveles de EPHB3 aumentan pero a medida que progresa el tumor hacia estadios más avanzados los niveles de expresión de esta proteína disminuyen, deteniendo la progresión tumoral (Cortina y Palomo-Ponce *et al*, 2007). PTK7 (*Protein tyrosin kinase 7*) pertenece a la familia de las protein-tirosina kinasas, que juegan un papel importante en la regulación de la proliferación y diferenciación celulares. Se ha descrito que PTK7 está sobreexpresada en células de cáncer colorrectal (Mossie *et al*, 1995). IGFBP3 (*Insulin-like growth factor binding*

protein-3) forma parte de un conjunto de proteínas que incluye los factores de crecimiento similares a insulina (*insulin-like*), sus receptores y las proteínas que se unen a ellos. Todas estas proteínas están relacionadas con la regulación de la proliferación y apoptosis celulares. IGFBP3 es la principal proteína de unión a IGF-1 (*Insulin-like growth factor 1*), niveles elevados de estas dos proteínas en suero se han relacionado con un mayor riesgo de padecer cáncer de mama (Renehan *et al*, 2004). TDGF1 (*Teratocarcinoma-derived growth factor 1*) es una proteína perteneciente a la familia EGF-CFC (*Epidermal growth factor- Cripto/FRL1/Criptic motif*), implicadas durante el desarrollo embrionario en la determinación del eje antero-posterior. Esta proteína no se detecta en tejidos de individuos adultos en condiciones normales, sin embargo, se ha visto altamente sobreexpresada en diferentes tipos de carcinomas epiteliales como colon, mama o páncreas (Salomon *et al*, 2000). La sobreexpresión de TDGF1 en tejidos tumorales se ha relacionado con una disminución en la expresión de E-cadherina, alterando la integridad del epitelio (Zhong *et al*, 2008).

Nuestro estudio muestra una buena correlación entre los resultados transcriptómicos y la validación proteómica en la mayoría de los genes analizados. La sobreexpresión fue especialmente significativa en el caso de ANXA3, LCN2, MMP7 , MMP11 y SPARC, que discriminan perfectamente entre tejido sano (niveles nulos de expresión) y tejido tumoral. Por otro lado, BMP4, LCN2 y MMP11 muestran una tendencia hacia los estadios tardíos del desarrollo tumoral. MMP11 actúa en la remodelación de la matriz extracelular por lo que podría tratarse de un buen marcador de estadios metastáticos de la neoplasia.

El patrón de tinción de algunos de las proteínas analizadas, presentes principalmente en el estroma tumoral, acentúa la necesidad de contemplar el tejido tumoral completo para la evaluación de las alteraciones presentes en un tumor porque algunas de dichas alteraciones son propias de las células del estroma que rodean a las tumorales y colaboran en la progresión tumoral (Mueller y Fusening, 2004).

Todas las dianas genómicas seleccionadas para su validación proteómica corresponden a genes con funciones muy interesantes para comprender el desarrollo de las neoplasias. Para analizar el efecto que los genes seleccionados tienen sobre líneas celulares de cáncer de colon estudiamos tanto los niveles de la expresión de los

genes seleccionados en las líneas celulares como las alteraciones que sufren las mismas al silenciar o aumentar la expresión de dichos genes.

Los análisis de expresión génica en ocho líneas celulares de cáncer de colon mostraron que tan sólo tres de los genes estudiados, ANXA3, BMP4 y MMP11, se sobreexpresan a nivel de RNA mensajero en comparación con muestras normales. Esto indica que las dianas obtenidas a partir de los análisis de *microarrays* de cDNA utilizando como material de partida tejidos tumorales reflejan tanto los genes alterados en las células tumorales propiamente dichas como las alteraciones presentes en el estroma que rodea la neoplasia.

Diferentes estudios evidencian la influencia del estroma tumoral en la progresión tumoral (Bissell y Radisky, 2001; Elenbaas y Weinberg, 2001). Durante la carcinogénesis, las células que forman el estroma tumoral juegan un papel importante por ser las responsables de elaborar un ambiente apropiado para la progresión de las células tumorales. El estudio de las interacciones entre las células tumorales y el estroma adyacente pone de manifiesto que las células no transformadas pertenecientes al estroma que rodea un tumor segregan sustancias que aumentan la capacidad tumorogénica de las células transformadas (Sneddon *et al*, 2006). Por otro lado, se ha demostrado que los fibroblastos pertenecientes al estroma tumoral presentan alteraciones genéticas diferentes de las alteraciones presentes en las células tumorales propiamente dichas (Patocs *et al*, 2007). Hay dos teorías sobre el origen de las alteraciones en las células del estroma celular. Una de las teorías postula que los fibroblastos que rodean a las células tumorales derivan de la misma célula progenitora que las células tumorales, como la progresión tumoral se basa en la acumulación progresiva de diferentes mutaciones génicas, las células que rodean al tumor podrían derivar de una célula transformada que no ha progresado hasta convertirse en tumoral; sin embargo, el hecho de que las mutaciones presentes en los fibroblastos no sean las mismas que las presentes en las células tumorales descarta esta teoría como origen de las mutaciones en los fibroblastos. La otra teoría postula que las células del estroma tumoral están expuestas a los daños causados por las especies reactivas de oxígeno (ROS) generadas por el alto metabolismo celular que tiene lugar en un tumor, estas ROS pueden dañar el DNA y causar mutaciones o deleciones que transformen una o varias células del estroma, que continuarán con su expansión clonal a la par que las

células tumorales. Este nuevo tipo de marcadores tumorales se ha utilizado para establecer una firma molecular propia del estroma tumoral para clasificar diferentes carcinomas primarios de mama (Bergamaschi *et al*, 2008) y constituyen un nuevo frente en la investigación de la progresión tumoral.

Se llevaron a cabo estudios funcionales de los genes ANXA3, BMP4 y MMP7 en las líneas celulares Sw480, Hct15 y Hct116 para determinar su influencia en la apoptosis y proliferación celulares. En el caso de BMP4, se observaron alteraciones tanto en su interferencia en la línea Sw480 como en su sobreexpresión en la línea Hct116. El papel de la vía de señalización mediada por las proteínas BMP durante la carcinogénesis es muy complejo dado que su sobreexpresión se ha relacionado tanto con la progresión como con la supresión tumoral. La principal evidencia de la relación de esta familia de proteínas en la progresión del cáncer colorrectal es el hecho de que la presencia de mutaciones en su receptor, MBPR1A, da lugar al síndrome Poliposis juvenil familiar, que aumenta la susceptibilidad a padecer cáncer colorrectal en un 12% (Howe *et al*, 2001). Las proteínas BMP se unen a su receptor, BMPR1, iniciando la cascada de señalización celular al fosforilar las proteínas Smad1, Smad5 y Smad8; a continuación, estas proteínas fosforiladas se asocian con Smad4 y se translocan al núcleo donde modelan la transcripción de los genes diana.

En el caso de Sw480, el silenciamiento de BMP4 dio lugar a un aumento en los niveles de apoptosis mientras que en la línea celular Hct116 su sobreexpresión redujo la proliferación celular. La línea Sw480 es deficiente en SMAD4, principal efector de la cascada de señalización mediada por las BMP, sin embargo, en nuestro estudio el silenciamiento de BMP4 en esta línea celular dio lugar a un aumento del 23% en los niveles de apoptosis, indicando que BMP4 puede inducir cambios en el crecimiento celular independientes de SMAD4. Confirmando este resultado, se ha descrito que esta línea celular presenta transcripción de los genes regulados por la vía BMP-SMAD a pesar de no expresar la proteína Smad4 (Beck *et al*, 2005).

En el caso de la línea celular Hct116 se observa que la sobreexpresión de BMP4 provoca un ligero descenso de la proliferación celular. Se ha descrito que la sobreexpresión de BMP4 en esta línea celular no afecta a la progresión celular sino que otorga un fenotipo invasivo a las células, cambiando su morfología e induciendo

la expresión de proteínas remodeladoras de la matriz extracelular que participan en la migración celular (Deng *et al*, 2007).

En resumen, durante el desarrollo de esta tesis doctoral se han validado a nivel tisular una serie de dianas genómicas identificadas como posibles marcadores de cáncer colorrectal. La progresión neoplásica de un tejido desde su estadio norma hasta el tumoral se caracteriza por una serie de alteraciones génicas que dan lugar a modificaciones en los patrones de expresión, función y estructura de las proteínas. La identificación de estos marcadores puede ser útil tanto a la hora de diagnosticar el cáncer en un estadio temprano como para disponer de nuevas dianas terapéuticas frente a las que dirigir fármacos.

Conclusiones

De los resultados obtenidos en esta tesis doctoral se puede concluir lo siguiente:

1. Se identificaron 337 genes alterados en cáncer colorrectal tras comparar la expresión génica diferencial entre tejido sano y tumoral utilizando *microarrays* de cDNA.
2. Se confirmó la alteración en los niveles de expresión génica por PCR cuantitativa sobre tejido tumoral y líneas celulares de cáncer de colon. Estos estudios reflejan que las dianas obtenidas no son sólo marcadores propios de las células tumorales del epitelio sino también marcadores del estroma que las rodea.
3. La obtención de clones de cDNA para la expresión de los productos proteicos con la secuencia apropiada constituye uno de los cuellos de botella a la hora de validar a nivel proteómico las alteraciones presentes a nivel genómico.
4. Se expresaron las proteínas en *E. coli* utilizando el sistema *Gateway*, muy eficiente para la expresión simultánea de múltiples proteínas. Las proteínas con un peso molecular superior a 50 kDa fueron prácticamente imposibles de obtener en *E. coli* utilizando esta estrategia. Por otro lado, MBP es la fusión más apropiada a la hora de obtener las proteínas solubles.
5. Se analizaron múltiples vías para la producción de anticuerpos. La tecnología del despliegue en fagos presenta problemas a la hora de obtener scFv solubles y funcionales en diagnóstico con una buena afinidad en un tiempo razonable.
6. Las librerías Mehta (no inmunes) fueron más eficaces que las Tomlinson (semisintéticas) a la hora de seleccionar anticuerpos humanos recombinantes debido a su mayor tamaño, se obtuvieron scFv con mayor afinidad y sin codones *stop*. Sin embargo, su aplicación en diagnóstico requiere su fusión a *tags* apropiados para su detección por técnicas inmunológicas, lo que ralentiza considerablemente su obtención y eficacia.
7. Se obtuvieron anticuerpos monoespecíficos frente a AGT, PTK7 y SPARC por la técnica PrEST. Sin embargo, no mostraron la especificidad y sensibilidad adecuadas para discernir entre tejido normal y tumoral mediante técnicas inmunológicas.
8. A pesar de requerir inmunización de animales, los anticuerpos mono y policlonales resultaron finalmente más rápidos, sensibles y eficientes en diagnóstico debido a su mayor afinidad por el antígeno.

9. Las dianas ANXA3, BMP4, CA9, LCN2, MMP7, MMP11 y SPARC ofrecieron una buena discriminación entre tejido sano y tumoral. Además, BMP4, LCN2 y MMP11 mostraron un mayor grado de expresión en los estadios tardíos del desarrollo tumoral, siendo su tinción negativa en tejidos normales e hiperplasias.
10. La proteína BMP4 ejerce funciones muy distintas dependiendo del tipo celular que se utilice como modelo. Su silenciamiento en la línea celular de cáncer de colon Sw480 promueve la apoptosis de las células por vías alternativas a la clásica mediada por SMAD mientras que su sobreexpresión en la línea Hct116 ralentiza la proliferación del cultivo celular.

Bibliografĩa

- Agaton, C., Falk, R., Höiden-Guthenberg, I., Göstring, L., Uhlén, M. and Hober, S. (2004) *Selective enrichment of monospecific polyclonal antibodies for antibody-based proteomics efforts*. J. Chromatogr. A. 1043, 33-40.
- Agaton, C., Galli, J., Guthenberg, I. H., Janzon, L., Hansson, M., Asplund, A., Brundell, E., Lindberg, S., Ruthberg, I., Wester, K., Wurtz, D., Höög, C., Lundeberg, J., Stahl, S., Pontén, F. and Uhlén, M. (2003) *Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues*. Mol. Cell. Proteomics 2, 405-414.
- Albala, J. S., Franke, K., McConnell, I. R., Pak, K. L., Folta, P. A., Karlak, B., Rubinfeld, B., Davies, A. H., Lennon, G. G. and Clark, R. *From genes to proteins: High-throughput expression and purification of the human proteome*. J. Cell. Biochem. 80, 187-191.
- Alessi, D. R., Sakamoto, K. and Bayascas, J. R. (2006) *LKB1-dependent signaling pathways*. Annu Rev Biochem 75, 137-163.
- Alfonso, P., Núñez, A., Madoz-Gúrpide, J., Lombardía, L., Sánchez, L. And Casal, J. I. (2004) *Proteomic expression analysis of colorectal cancer by two-dimensional differential gel electrophoresis*. Proteomics 5, 2602-2611.
- Alitalo, K., Schwab, M., Lin, C. C., Varmus, H. and Bishop, J. M. (1983) *Homogeneously stained chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma*. Proc. Natl. Acad. Sci. U.S.A. 80, 1707-1711.

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic Acids Res. 25, 3389-3402.
- Arnold, S. F., Tims, E. and McGrath, B. E. (1999) *Identification of bone morphogenetic proteins and their receptors in human breast cancer cell lines: importance of BMP2*. Cytokine 11, 1031-1037.
- Ashkar, S., Weber, G. F., Panoutsakopoulou, V., Sanchirico, M. E., Jansson, M., Zawaideh, S., Rittling, S. R., Denhardt, D. T., Glimcher, M. J. and Cantor, H. (2000) *Eta-1 (osteopontin): an early component of type-1 (cell-mediated) immunity*. Science 287, 860-864.
- Aslan, K., Lakowicz, J. R. and Geddes, C. (2005) *Plasmon light scattering in biology and medicine: new sensing approaches, visions and perspectives*. Curr. Opin. Chem. Biol. 9, 538-544.
- Bach, S. P., Renehan, A. G. and Potten, C. S. (2000) *Stem cells: the intestinal stem cell as a paradigm*. Carcinogenesis 21, 469-476.
- Balemans, W. and Van Hul, W. (2002) *Extracellular regulation of BMP signalling in vertebrates: a cocktail of modulators*. Dev. Biol. 250, 231-250.
- Barbas, C., Kang, A.S., Lerner, R.A. and Benkovic, S.J. (1991) *Assembly of combinatorial antibody libraries on phage surfaces: the gene III site*. Proc. Natl. Acad. Sci. U.S.A. 88, 7978-7982.
- Barderas, R., Shochat, S., Martínez-Torrecuadrada, J. L., Altschuh, D., Meloen, R. And Casal, J. I. (2006) *A fast mutagenesis procedure to recover soluble and functional scFvs containing amber stop codons from synthetic and semisynthetic antibody libraries*. J. Immunol. Methods 312, 182-189.

- Bast, R. C. Jr., Feeney, M., Lazarus, H., Nadler, L. M., Colvin, R. B. and Kanapp, R. C. (1981) *Reactivity of a monoclonal antibody with human ovarian carcinoma*. J. Clin. Invest. 68, 1331-1337.
- Battle, E., Henderson, J. T., Beghtel, H., van den Born, M. M. W., Sancho, E., Huls, G., Meeldijk, J., Robertson, J., van de Wetering, M., Pawson, T. and Clevers, H. (2002) *β -Catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/EphrinB*. Cell 111, 251-263.
- Beck, S. E., Jung, B. H., Fiorino, A., Gómez, J., Del Rosario, E., Cabrera, B. L., Huang, S. C., Chow, J. Y. C. and Carethers, J. M. (2005) *Bone morphogenetic protein signaling and growth suppression in colon cancer*. Am. J. Gastrointest. Liver Physiol. 291, 135-145.
- Begent, R. H., Verhaar, M. J., Chester, K. A., Casey, J. L., Green, A. J., Napier, M. P., Hope-Stone, L. D., Cushen, N., Keep, P. A., Johnson, C. J., Hawkins, R. E., Hilson, A. J. and Robson, L. (1996) *Clinical evidence of efficient tumor targeting based on single-chain Fv antibody selected from a combinatorial library*. Nat. Med. 9, 979-984.
- Bergamaschi, A., Tagliabue, E., Sorlie, T., Naume, B., Triulzi, T., Orlando, R., Russnes, H. G., Nesland, J. M., Tammi, R., Auvinen, P., Kosma, V. M., Mènard, S. and Borresen-Dale, A. L. (2008) *Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome*. J. Pathol. 214, 357-367.
- Bhanot, P., Brink, M., Samos, C. H., Hsieh, J. C., Wang, Y., Macke, J. P., Andrew, D., Nathans, J. and Nusse, R. (1996) *A new member of the frizzled family from Drosophila functions as a Wingless receptor*. Nature 382, 225-230.

- Bissell, M. J. and Radisky, D. (2001) *Putting tumours in context*. Nat. Rev. Cancer 1, 46-54.
- Boulianne, G. L., Hozumi, N. and Shulman, M. J. (1984) *Production of functional chimaeric mouse/human antibody*. Nature 321, 643-646.
- Bos, J. L. (1989) *ras oncogenes in human cancer: a review*. Cancer Res. 49, 4682-4689.
- Bradshaw, A. D., Graves, D. C., Motamed, K. and Sage, E. H. (2003) *SPARC-null mice exhibit increased adiposity without significant differences in overall body weight*. Proc. Natl. Acad. Sci. U.S.A. 100, 6045-6050.
- Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E. and LaBaer, J. (2002) *Proteome-scale purification of human proteins from bacteria*. Proc. Natl. Acad. Sci. U.S.A. 99, 2654-2659.
- Breitling, S.D., Seehaus, T., Klewinghaus, I. And Little, M. (1991) *A surface expression vector for antibody screening*. Gene 104, 147-153.
- Brizuela, L., Richardson, A., Marsischky, G. and LaBaer, J. (2002) *The FLEXGene repository: exploiting the fruits of the genome projects by creating a needed resource to face the challenges of the post-genomic era*. Arch. Med. Res. 33, 318-324.
- Cañas, B., López-Ferrer, D. Ramos-Fernández, A., Camafeita, E. and Calvo, E. (2006) *Mass spectrometry Technologies for proteomics*. Brief Funct. Genomic Proteomic 4, 295-320.
- Carter, P. J. (2006) *Potent antibody therapeutics by design*. Nat. Rev. Immunol. 6, 343-357.
- Célérier, J., Cruz, A., Lamandé, N., Gasc, J. M. and Corvol, P. (2002) *Angiotensinogen and its cleaved derivatives inhibit angiogenesis*. Hypertension 39, 224-228.

- Chakraborty, G., Jain, S. and Kundu, G. C. (2008) *Osteopontin promotes vascular endothelial growth factor-dependent breast tumor growth and angiogenesis via autocrine and paracrine mechanisms*. Cancer Res. 68, 152-161.
- Chlenski, A., Liu, S., Guerrero, I. J., Yang, Q., Tian, Y., Salwen, H. R., Zage, P. And Cohn, S. L. (2006) *SPARC expression is associated with impaired tumor growth, inhibited angiogenesis and changes in the extracellular matrix*. Int. J. Cancer 118, 310-316.
- Cleven, A. H., van Engeland, M., Wouters, B. G. and de Bruine, A. P. (2007) *Stromal expression of hypoxia regulated proteins is an adverse prognostic factor in colorectal carcinomas*. Cell Oncol. 29, 229-240.
- Cortina, C., Palomo-Ponce, S., Iglesias, M., Fernández-Masip, J. L., Vivancos, A., Whissell, G., Humà, M., Peiró, N., Gallego, L., Jonkheer, S., Davy, A., Lloreta, J., Sancho, E. and Batlle, E. (2007) *EphB-ephrin-B interactions suppress colorectal cancer progression by compartmentalizing tumor cells*. Nat. Genet. 39, 1376-1383.
- Cowland, J. B. and Borregaard, N. (1997) *Molecular characterization and pattern of tissue expression of the gene for neutrophil gelatinase-associated lipocalin from humans*. Genomics 45, 17-23.
- Deng, H., Makizumi, R., Ravikumar, T. S. and Dong, H. (2007) *Bone morphogenetic protein-4 is overexpressed in colonic adenocarcinomas and promotes migration and invasion of HCT116 cells*. Exp. Cell Res. 313, 1033-1044.
- DeRisi, J. Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A. and Trent, J. M. (1996) *Use of a cDNA microarray to analyse gene expression patterns in human cancer*. Nat. Genet. 14, 457-460.

- Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A. and Chinnaiyan, A. M. (2001) *Delineation of prognostic biomarkers in prostate cancer*. Nature 412, 822-826.
- Duenas, M. and Borrebaeck, C. A. (1995) *Novel helper phage design: intergenic region affects the assembly of bacteriophages and the size of antibody libraries*. FEMS Microbiol. Letts. 125, 317-321.
- Dukes, C. E. (1932) *The classification of cancer in the rectum*. J. Pathol. Bacteriol 35, 323.
- Elenbaas, B. and Weinberg, R. A. *Heterotypic signaling between epithelial tumor cells and fibroblasts in carcinoma formation*. Exp. Cell Res. 264, 169-184.
- Fearon, E. C., Hamilton, S. R. and Vogelstein, B. (1987) *Clonal analysis of human colorectal tumors*. Science 238, 193- 197.
- Fernández, L. A. (2004) *Prokaryotic expression of antibodies and affibodies*. Curr. Opin. Biotechnol. 15, 364-373.
- Foote, J. and Eisen, H. N. (1995) *Kinetic and affinity limits on antibodies produced during immune responses*. Proc. Natl. Acad. Sci. U.S.A. 92, 1254-1256.
- Friedl, A., Stoesz, S. P., Buckley, P. and Gould, M. N. (1999) *Neutrophil gelatinase-associated lipocalin in normal and neoplastic human tissues. Cell type-specific pattern of expression*. Histochemistry Journal 31, 433-441.
- Glockshuber, R., Schmidt, T. and Plückthun, A. (1992) *The disulfide bonds in antibody variable domains: effects on stability, folding in vitro and functional expression in Escherichia coli*. Biochemistry 31, 1270-1279.
- Gold, P. and Freedman, S. O. (1975) *Test for carcinoembryonic antigen. Role in diagnosis and management of cancer*. JAMA 234, 190-192.

- Goldblum, S. E., Ding, X., Funk, S. E. and Sage, E. H. (1994) *SPARC (secreted protein acidic and rich in cysteine) regulates endothelial cell shape and barrier function*. Proc. Nat. Acad. Sci. U.S.A. 91, 3448-3452.
- Griffiths, A. D., Malmqvist, M., Marks, J. D., Bye, J. M., embleton, M. J. and McCafferty, J., Baier, M., Holliger, K. P., Gorick, B. D., Hughes-Jones, N., Hoogenboom, H., R. and Winter, G. (1993) *Human anti-self antibodies with high specificity from phage display libraries*. EMBO J. 12, 1203-1210.
- Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L. and Robertson, M. (1991) *Identification and characterization of the familial adenomatous polyposis coli gene*. Cell 66, 589-600.
- Hamada, S., Satoh, K., Hirota, M., Kimura, K., Kanno, A., Masamune, A., and Shimosegawa, T. (2007) *Bone morphogenetic protein 4 induces epithelial-mesenchymal transition through MSX2 induction on pancreatic cancer cell line*. J. Cell. Physiol. 213, 768-774.
- Hammarström, M., Hellgren, N., Van Den Berg, S., Berglund, H. and Härd, T. (2002) *Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli*. Protein Sci. 11, 313-321.
- Hanson, C. A. and Miller, J., R. (2005) *Non-traditional roles for the Adenomatous Polyposis Coli (APC) tumor suppressor protein*. Gene 361, 1-12.
- Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998) *SOSUI: classification and secondary structure prediction system for membrane proteins*. Bioinformatics 14, 378-379.

- Hirose, M., Hayano, T., Shirai, H., Nakamura, H. and Kikuchi, M. (1998) *Isolation of anti-glutathione antibodies from a phage display library*. Protein Eng. 11, 243-248.
- Hoogenboom, H.R. *Antibody Phage Display: Methods and Protocols*. Methods in Molecular Biology, vol. 178.
- Hoogenboom, H.R., Griffiths, A.D., Johnson, K.S., Chiswell, D.J., Hudson, P. and Winter, G. (1991) *Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains*. Nucleic Acids Res. 19, 4133-4137.
- Howe, J. R., Bair, J. L., Sayed, M. G., Anderson, M. E., Mitros, F. A., Petersen, G. M., Velculescu, V. E., Traverso, G. and Vogelstein, B. (2001) *Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis*. Nat. Genet. 28, 184-187.
- Hunakova, L., Bodo, J., Chovancova, J., Sulikova, G., Pastorekova, S. and Sedlak, J. (2007) *Expresión of new prognostic markers, peripheral-type benzodiazepine receptor and carbonic anhydrase IX, in human breast and ovarian carcinoma cell lines*. Neoplasma 54, 541-548.
- Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. and Perucho, M. (1993) *Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis*. Nature 363, 558-561.
- Järvelä, S., Parkkila, S., Bragge, H., Kähkönen, M., Parkkila, A. K., Soini, Y., Pastorekova, S., Pastorek, J. and Haapasalo, H. (2008) *Carbonic anhydrase IX in oligodendroglial brain tumors*. BMC Cancer 9, R89.

- Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. and Winter, G. (1986) *Replacing the complementarity-determining regions in a human antibody with those from a mouse*. Nature 321, 522-525.
- Kehoe, J. W. and Kay, B. K. (2005) *Filamentous phage display in the new millennium*. Chem. Rev. 105, 4056-4072.
- Kiian, I., Tkachuk, N., Haller, H. and Dumler, I. (2003) *Urokinase-induced migration of human vascular smooth muscle cells requires coupling of the small GTPases RhoA and Rac1 to the Tyk2/PI3-K signalling pathway*. Thromb. Haemost. 89, 904-914.
- Kim, J. H., Skates, S. J., Wong, K., Schorge, J. O., Felmate, C. M., Berkowitz, R. S., Cramer, D. W. and Mock, S. C. (2002) *Osteopontin as a potential biomarker for ovarian cancer*. The journal of the American Medical Association 287, 1671-1679.
- Kim, J. S., Crooks, H., Dracheva, T., Nishanian, T. G., Singh, B., Jen, J. and Waldman, T. (2002) *Oncogenic b-catenin is required for bone morphogenetic protein 4 expression in human cancer cells*. Cancer Res 62, 2744-2748.
- Klein, J. and Dhurjati, P. (1995) *Protein aggregation kinetics in an Escherichia coli strain overexpressing a Salmonella typhimurium CheY mutant gene*. Appl. Environ. Microbiol. 61, 1220-1225.
- Köhler, G. and Milstein, C. (1975) *Continuous cultures of fused cells secreting antibody of predefined specificity*. Nature 256, 495-497.
- Komarova, N. L., Sengupta, A. and Nowak, M. A. (2003) *Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability*. Journal of Theoretical Biology 223, 433-450.

- Korinek, V., Barker, N., Morin, P. J., van Wichen, D., de Weger, R., Kinzler, K. W., Vogelstein, B. and Clevers, H. (1997) *Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC^{-/-} colon carcinoma*. Science 275, 1784-1787.
- Kristensen, P. and Winter, G. (1998) *Proteolytic selection for protein folding using filamentous bacteriophages*. Fold. Des. 3, 321-328.
- Kyte, J. and Doolittle, R. (1982) *A simple method for displaying the hidropathic character of a protein*. Journal of Molecular Biology 157, 105-132.
- Laemmli, U. K. (1970) *Most commonly used discontinuous buffer system for SDS electrophoresis*. Nature 227, 680-686.
- Landy, A. (1989) *Dynamic, Structural, and Regulatory Aspects of lambda Site-Specific Recombination*. Ann Rev. Biochem 58, 913-941.
- Leath 3rd, C. A., Douglas, J. T., Curiel, D. T. and Álvarez, R. D. (2004) *Single-chain antibodies: A therapeutic modality for cancer gene therapy*. Int. J. Oncol. 24, 765-771.
- Lengauer, C., Kinzler, K. and Vogelstein, B. (1998) *Genetic instabilities in human cancers*. Nature 396, 643-649.
- Lennon, G. G., Auffray, C., Polymeropoulos, M. and Soares, M. B. *The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression*. (1996) Genomics 33, 151-152.
- Lin, Y., Furukawa, Y., Tsunoda, T., Yue, C. T., Yang, K. C. and Nakamura, Y. (2002) *Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas*. Oncogene 21, 4120-4128.

- Lindskog, M., Rockberg, J., Uhlén, M. and Sterky, F. (2005) *Selection of protein epitopes for antibody production*. BioTechniques 38, 723-727.
- Liu, Z. X., Yi, G. H., Qi, Y. P., Liu, Y. L., Yan, J. P., Qian, J., Du, E. Q. and Ling, W. F. (2005) *Identification of single-chain antibody fragments specific against SARS-associated coronavirus from phage-display antibody library*. Biochem. Res. Commun. 329, 437.
- Lowe, D. and Jermutus, L. (2004) *Combinatorial protein biochemistry for therapeutics and proteomics*. Curr. Pharm. Biotechnol. 5, 17-27.
- Lynch, P. M., Wargovich, M. J., Lynch, H. T., Palmer, C., Lanspa, S., Drouhard, T. and Lynch, J. (1991) *A follow-up study of colonic epithelial proliferation as a biomarker in a native American family with hereditary nonpolyposis colon cancer*. J. Natl. Cancer Inst. (Bethesda) 83, 951-954.
- Madoz-Gúrpide, J., Cañamero, M., Sánchez, L., Solano, J., Alfonso, P. And Casal, J. I. (2007) *A Proteomic analysis of cell signaling alterations in colorectal cancer*. Mol. Cell. Proteomics 12, 2150-2164.
- Madoz-Gúrpide, J., López-Serra, P., Martínez-Torrecuadrada, J. L., Sánchez, L., Lombardía, L. and Casal, I. (2006) *Proteomics-based Validation of Genomic Data*. Mol. Cell. Proteomics 8, 1471-1483.
- Mandel, J. S., Church, T. R., Ederer, F. and Bond, J. H. (1999) *Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood*. J. Natl. Cancer Inst. 91, 434-437.
- Marouga, R., David, S. and Hawkins, E. (2005) *The development of the DIGE system: 2D fluorescence difference gel analysis technology*. Anal. Bioanal. Chem. 382, 669-678.

- Martin, C. D., Rojas, G., Mitchell, J. N., Vincent, K., J., Wu, J., McCafferty, J. and Schofield, D. J. (2006) *A simple vector system to improve performance and utilisation of recombinant antibodies*. BioMed Central Biotechnology 6:46.
- Martínez, N., Camacho, F. I., Algara, P., Rodríguez, A., Dopazo, A., Ruiz-Ballesteros, E., Martin, P., Martínez-Climent, J. A., García-Conde, J., Menarguez, J., Solano, F., Mollejo, M. and Piris, M. A. (2003) *The molecular signature of mantle cell lymphoma reveals multiple signals favoring cell survival*. Cancer Res. 63, 8226-8232.
- Martínez-Torrecuadrada, J. L., Romero, S., Nuñez, A., Alfonso, P., Sánchez-Céspedes, M. And Casal, J. I. (2005) *An efficient expression system for the production of functionally active human LKBI*. Journal of Biotechnology 115, 23-34.
- Martínez-Torrecuadrada, J. L., Cifuentes, G., López-Serra, P., Saenz, P., Martínez, A. and Casal, J. I. (2005) *Targeting the extracellular domain of fibroblast growth factor receptor 3 with human single-chain Fv antibodies inhibits bladder carcinoma cell line proliferation*. Clin. Cancer Res. 11, 6280-6290.
- Marsischky, G. and LaBaer, J. (2004) *Many paths to many clones: a comparative look at high-throughput cloning methods*. Genome Res. 14, 2020-2028.
- Masaki, T., Matsuoka, H., Sugiyama, M., Abe, N., Goto, A., Sakamoto, A. and Atomi, Y. (2001) *Matrylisin (MMP-7) as a significant determinant of malignant potential of early invasive colorectal carcinomas*. Brit. J. Cancer 84, 1317-1321.
- Mason, D. Y., Cordell, J. L. and Pulford K.A.F. (1983) *Production of monoclonal antibodies for immunocytochemical use*. In Bullock GR, Petrusz P, eds. Techniques in Immunocytochemistry. London, Academic Press, 175-216.

- McCafferty, J., Griffiths, A. D., Winter, G. and Chiswell, D. J. (1990) *Phage antibodies: filamentous phage displaying antibody variable domains*. Nature 348, 552-554.
- Minn, A. J., Gupta, G. P., Siegel, P. M., Bos, P. D., Shu, W., Giri, D. D., Viale, A., Olshen, A. B., Gerald, W. L. and Massagué, J. (2005) *Genes that mediate breast cancer metastasis to lung*. Nature 436, 518-524.
- Morimoto, I., Sasaki, Y., Ishida, S., Imai, K. and Tokino, T. *Identification of the osteopontin gene as a direct target of TP53*. Genes Chromosomes Cancer 33, 270-278.
- Morrison, S. L., Johnson, M. J., Herzenberg, L. A. and Oi, V. T. (1984) *Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains*. Proc. Natl. Acad. Sci. U.S.A. 81, 6851-6855.
- Mossie, K., Jallal, B., Alves, F., Sures, I., Plowman, G. D. and Ullrich, A. (1995) *Colon carcinoma kinase-4 defines a new subclass of the receptor tyrosine kinase family*. Oncogene 16, 2179-2184.
- Mueller, M. M. and Fusening, N. E. (2004) *Friends or foes: bipolar effects of the tumour stroma in cancer*. Nat. Rev. Cancer 4, 839-849.
- Nielsen, B. S., Rank, F., Illemann, M., Lund, L. R. and Dano, K. (2007) *Stromal cells associated with early invasive foci in human mammary ductal carcinoma in situ coexpress urokinase and urokinase receptor*. Int. J. Cancer 120, 2086-2095.
- Pastorekova, S., Zatovicova, M. and Pastorek, J. (2008) *Cancer-associated carbonic anhydrases and their inhibition*. Curr. Pharm. Des. 14, 685-698.

- Patocs, A., Zhang, L., Xu, Y., Weber, F., Caldes, T., Mutter, G. L., Platzer, P. and Eng, C. (2007) *Breast-cancer stromal cells with TP53 mutations and nodal metastases*. N. Engl. J. Med. 357 (25), 2537-2538.
- Pinson, K. I., Brennan, J., Monkley, S., Avery, B. J. and Skarnes, W. C. (2000) *An LDL-receptor-related protein mediates Wnt signaling in mice*. Nature 407, 535-538.
- Polakis, P. (1999) *The oncogenic activation of b-catenin*. Curr. Opin. Genet. Dev. 9, 15-21.
- Possee, R. D., Thomas, C. J. and King, L. A. (1999) *The use of baculovirus vectors for the production of membrane proteins in insect cells*. Biochemical Society Transactions 27 (6) 928-932.
- Rajagopalan, H., Nowak, M. A., Vogelstein, B. and Lengauer, C. (2003) *The significance of unstable chromosomes in colorectal cancer*. Nat. Rev. Cancer 3, 695-700.
- Riechmann, L., Clark, M., Waldmann, H. and Winter, G. (1988) *Reshaping human antibodies for therapy*. Nature 332, 323-327.
- Reiter, Y., Brinkmann, U., Kreitman, R. J., Jung, S. H., Lee, B. and Pastan, I. (1994) *Stabilization of the Fv fragments in recombinant immunotoxins by disulfide bonds engineered into conserved framework regions*. Biochemistry 33, 5451-5459.
- Renehan, A.G., Zwahlen, M., Minder, C., O'Dwyer, S. T., Shalet, S. M. and Egger, M. (2004) *Insulin-like growth factor (IGF-1), IGF binding protein-3 and cancer risk: systematic review and meta-regression analysis*. Lancet 363, 1346-1353.

- Roncador, G., García, J. F., Maestre, L., Lucas, E., Menarguez, J., Ohshima, K., Nakamura, S., Banham, A. H. and Piris, M. A. (2005) *FOXP3, a selective marker for a subset of adult T-cell leukaemia/lymphoma*. Leukemia 19, 2247-2253.
- Rothhammer, T., Poser, I., Soncin, F., Bataille, F., Moser, M. and Bosserhoff, A., K. (2005) *Bone morphogenetic proteins are overexpressed in malignant melanoma and promote cell invasion and migration*. Cancer Res. 65, 448-456.
- Sachdev, D. and Chirgwin, J. M. (1998) *Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin*. Protein Expr. Purif. 12, 122-132.
- Sadoshima, J., Xu, Y., Slayter, H. S. and Izumo, S. (1993) *Autocrine release of angiotensin II mediates stretch-induced hypertrophy of cardiac myocytes in vitro*. Cell 75, 977-84.
- Salomon, D. S., Bianco, C., Ebert, A. D., Khan, N. I., Se Santis, M., Normando, N., Wechselberger, C., Seno, M., Williams, K., Sanicola, M., Foley, S., Gullick, W. J. and Persico, G. (2000) *The EGF-CFC family: Novel epidermal growth factor-related proteins in development and cancer*. Endocr. Relat. Cancer 7, 199-226.
- Schottenfeld, D., Gantt, R. C. and Wyner, E. L. (1974) *The role of alcohol and tobacco in multiple primary cancers of the upper digestive system, larynx and lung: a prospective study*. Prev. Med. 3, 277-293.
- Schultz, J., Milpetz, F., Bork, P. and Ponting, C. P. (1998) *SMART, a simple modular architecture research tool: identification of signalling domains*. Proc. Natl. Acad. Sci. U.S.A. 95, 5857-5864.

- Shalon, D., Smith, S. J. and Brown, P. O. (1996) *A DNA Microarray System for Analyzing Complex DNA Samples Using Two-color Fluorescent Probe Hybridization*. *Genome Research* 6, 639-645.
- Sharpe C. R, Siemiatycki J. A, Rachet B. P. (2002) *The effects of smoking on the risk of colorectal cancer*. *Dis colon rectum*. 45, 1041-50.
- Shaw, R. J. and Cantley, L. C. (2006) *Ras, PI(3)K and mTOR signaling controls tumor cell growth*. *Nature* 441, 424-430.
- Skoglund, J., Emterling, A., Arbman, G., Anglard, P. and Sun, X. F. (2004) *Clinicopathological significance of stromelysin-3 expression in colorectal cancer*. *Oncology* 67, 67-72.
- Slattery, M. L., Anderson, K., Curtin, K., Ma, K., Schaffer, D., Edwards, S. and Samowitz, W. (2001) *Lifestyle factors and Ki-ras mutations in colon cancer tumors*. *Mutat. Res.* 483, 73-81.
- Smith, G.P. (1985) *Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface*. *Science* 228, 1315-1317.
- Smyth, G. K. and Speed, T. (2003) *Normalization of cDNA microarray data*. *Methods* 31, 265-273.
- Sneddon, J.B., Hanson, H. Z., Montgomery, K., van de Rijn, M., Tward, A. D., West, R., Gladstone, H., Chang, H. Y., Morganroth, G. S., Oro, A. E., Brown, P. O. (2006) *Bone morphogenetic protein antagonist gremlin 1 is widely expressed by cancer-associated stromal cells and can promote tumor cell proliferation*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14842-14847.
- Söreide, K., Janssen, E. A. M., Söiland, H., Körner, H. and Baak, J. P. A. (2006) *Microsatellite instability in colorectal cancer*. *Br. J. Surg.* 93, 395-406.

- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Lonning, P. E. and Borresen-Dale, A. (2001) *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications*. Proc. Natl. Acad. Sci. U.S.A. 98, 10869-10874.
- Stamey, T. A., Yang, N., Hay, A. R., McNeal, J. E., Freiha, F. S. and Redwine, E. (1987) *Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate*. N. Engl. J. Med. 317, 909-916.
- Strieter, R. M., Polverini, P. J., Kunkel, S. L., Arenberg, D. A., Burdick, M. D., Kasper, J., Dzuiba, J., Van Damme, J., Walz, A., Marriott, D., Chan, S. Y., Roczniak, S. and Shanafelt, A. B. (1995) *The functional role of the ELR motif in CXC chemokine-mediated angiogenesis*. J. Biol. Chem. 270, 27348-27357.
- Sweet, K., Willis, J., Zhou, X. P., Gallione, C., Sawada, T., Alhopuro, P., Khoo, S. K., Patocs, A., Martin, C., Bridgeman, S., Heinz, J., Pilarski, R., Lehtonen, R., Prior, T. W., Fredborg, T., The, B. T., Marchuk, D. A., Aaltonen, L. A. and Eng, C. (2005) *Molecular classification of patients with unexplained hamartomatous and hyperplastic polyposis*. JAMA 294, 2465-2473.
- Tang, H., Wang, J. Bai, F., Zhai, H., Gao, J., Hong, L., Xie, H., Zhang, F., Lan, M., Yao, W., Liu, J., Wu, K. and Fan, D. (2008) *Positive correlation of osteopontin, cyclooxygenase-2 and vascular endothelial growth factor in gastric cancer*. Cancer Invest. 26, 60-67.

- Thykjaer, T., Workman, C., Kruhoffer, M., Demtroder, K., Wolf, H., Andersen, L. D., Frederiksen, C. M., Knudsen, S. and Orntoft, T. F. (2001) *Identification of gene expression patterns in superficial and invasive human bladder cancer*. Cancer Res. 61, 2492-2499.
- Tracey, L., Villuendas, R., Ortiz, P., Dopazo, A., Spiteri, I., Lombardía, L., Rodríguez-Peralto, J. L., Fernández-Herrera, J., Hernández, A., Fraga, J., Domínguez, O., Herrero, J., Alonso, M. A., Dopazo, J. and Piris, M. A. (2002) *Identification of genes involved in resistance to interferon- α in cutaneous T-cell lymphoma*. Am. J. Pathol. 161, 1825-1837.
- Tusher, V. G., Tibshirani, R. and Chhu, G. (2001) *Significance analysis of microarrays applied to the ionizing radiation response*. Proc. Natl. Acad. Sci. U.S.A. 98, 5116-5121.
- Uhlén, M., Björling, E., Agaton, C. et al (2005) *A human protein atlas for normal and cancer tissues based on antibody proteomics*. Mol. Cell. Proteomics 4.12, 1920-1932.
- Uhlén, M. and Pontén, F. (2005) *Antibody-based proteomics for human tissue profiling*. Mol. Cell. Proteomics 4.4, 384-393.
- van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A. P., Tjon-Pon-Fong, M., Moerer, P., van den Born, M., Soete, G., Pals, S., Eilers, M., Medema, R. and Clevers, H. (2002) *The b-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells*. Cell 111, 241-250.

- Verhoeyen, M., Milstein, C. and Winter, G. (1988) *Reshaping human antibodies: grafting an antilysozyme activity*. Science 239, 1534-1536.
- Waite, K. A. and Eng, C. (2002) *Protean PTEN: form and function*. Am J Hum Genet 70, 829-844.
- Walhout, A. J., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., van den Heuvel, S. and Vidal, M. (2000) *Gateway recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes*. Methods Enzymol. 328, 575-592.
- Woerner S. M., Benner, A., Sutter, C., Schiller, M., Yuan, Y. P., Keller, G., Bork, P., von Knebel Doeberitz, M and Gebert, J. F. (2003) *Pathogenesis of DNA repair-deficient cancers: a statistical meta-analysis of putative Real Common Target genes*. Oncogene 22, 2226-2253.
- Xu, J. L. and Davis, M. M. (2000) *Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities*. Immunity 13, 37-45.
- Yan, J. P., Ko, J. H. and Qi, Y. P. (2004) *Generation and characterization of a novel single-chain antibody fragment specific against human fibrin clots from phage display antibody library*. Throm. Res. 114, 205-211.
- Yau, K. Y., Dubuc, G., Li, S., Hiramata, T., Mackenzie, C. R., Jermutus, L., Hall, J. C. and Tanha, J. (2005) *Affinity maturation of a V(H)H by mutational hotspot randomization*. J. Immunol. Methods 297, 213-224.
- Zeng, Z., S., Shu, W. P., Cohen, A. M. and Guillem, J. G. (2002) *Matrix metalloproteinase-7 expression in colorectal cancer liver metastases: evidence for involvement of MMP-7 activation in human cancer metastases*. Clin. Cancer Res. 8, 144-148.

Zhong, X. Y., Zhang, L. H., Jia, S. Q., Shi, T., Niu, Z. J., Du, H., Zhang, G. G., Hu, Y., Lu, A. P., Li, J. Y. and Ji, J. F. (2008) *Positive association of up-regulated Cripto-1 and down-regulated E-cadherin with tumour progression and poor prognosis in gastric cancer*. Histopathology 52, 560-568.

Anexo I

Table S1. Differentially expressed genes of twenty two paired samples of CRC patients, detected by one class SAM analysis at a delta value threshold corresponding to a q-value<1%

Symbol	GenBank Accession	q-value (%)	Biopsy#9010085_56T/55N_B	Biopsy#6010533_46T/45N_C	Biopsy#9010070_54T/53N_C	Biopsy#5011392_48T/47N_B	Biopsy#9010086_58T/57N_C	Biopsy#6010532_40T/39N_A	Biopsy#1010118_36T/35N_C	Biopsy#1010283_80T/78N_C	Biopsy#1010445_83T/85N_C	Biopsy#3011002_1T/15N_B	Biopsy#1010320_77T/90N_C	Biopsy#1010322_96T/95N_C	Biopsy#1010433_84T/87N_C	Biopsy#1010286_92T/71N_C	Biopsy#3011021_30T/29N_B	Biopsy#3011003_59T/37N_B	Biopsy#1010414_107T/106N_C	Biopsy#1010455_109T/108N_C	Biopsy#5011390_62T/41N_A	Biopsy#1010292_81T/72N_C	Biopsy#3011011_20T/26N_A	Biopsy#1010457_111T/110N_C
MMP3	W51794	0.000	2.3	1.2	2.1	3.7	1.5	0.7	1.3	NA	4.2	2.6	1.0	2.2	1.4	2.5	2.7	NA	6.7	NA	2.4	2.1	5.0	4.7
THBS2	H38240	0.000	3.7	0.0	1.5	2.0	1.9	3.5	0.9	3.3	2.0	NA	3.9	2.0	-1.2	NA	3.9	NA	3.3	2.4	NA	3.8	7.7	0.7
CTHRC1	AA410434	0.000	2.6	0.8	2.4	3.0	2.1	1.8	0.0	3.1	3.6	0.8	3.4	3.0	-0.4	-0.1	4.7	2.5	4.6	2.3	3.7	3.6	5.7	0.9
PLAU	AA284668	0.000	2.2	1.9	1.8	2.1	0.8	1.9	1.5	2.4	2.2	0.5	2.4	1.7	1.1	1.6	1.9	4.3	9.1	NA	4.1	2.1	3.1	1.8
CXCL1	W42723	0.000	1.6	3.5	4.0	NA	0.8	3.3	2.1	2.4	2.7	3.0	1.8	1.1	0.8	2.8	2.2	3.4	3.5	0.5	2.3	3.7	1.2	2.2
CXCL3	AA935273	0.000	1.2	3.7	3.4	2.2	0.9	3.6	1.5	1.8	3.5	2.4	2.2	1.4	0.9	1.9	2.4	3.8	2.3	0.1	1.9	4.1	1.9	2.2
CXCL1	W46900	0.000	2.3	3.5	3.8	2.1	0.9	4.4	2.1	2.4	2.9	2.3	1.7	1.1	0.4	3.2	1.2	3.2	NA	-0.1	2.0	3.5	1.5	2.1
TIMP1	BE617255	0.000	3.6	NA	1.2	2.3	2.5	2.8	1.0	2.2	1.6	1.8	2.2	1.9	-0.5	2.0	0.7	1.5	4.6	0.7	2.3	1.7	3.3	3.9
MMP11	AA954935	0.000	0.5	-0.1	2.3	3.3	2.3	1.9	0.6	2.4	1.8	0.1	1.4	2.8	0.6	1.1	NA	3.4	5.2	2.1	4.2	2.9	3.9	0.2
SULF1	AI653116	0.000	2.3	1.0	1.3	2.0	2.3	2.2	0.4	NA	2.4	0.7	3.3	1.5	0.7	0.9	2.6	1.8	3.8	0.2	3.3	3.7	NA	2.3
SPP1	AA775616	0.000	2.3	0.3	1.0	1.5	1.4	1.1	0.0	2.7	1.8	0.9	5.1	2.0	0.8	0.4	4.6	2.6	3.4	0.0	2.4	1.4	3.1	3.1
MMP12	R92994	0.000	2.6	1.9	3.8	2.1	-0.4	1.7	0.8	2.6	2.8	1.9	1.7	-0.2	1.8	2.0	1.9	1.8	2.4	0.6	1.7	2.0	3.0	2.9
COL12A1	AA478481	0.000	2.7	1.1	1.5	1.7	2.0	1.4	0.9	2.9	1.6	0.1	3.3	1.1	1.0	0.8	2.2	2.3	2.9	1.4	2.6	3.1	3.1	1.6
LCN2	AA401137	0.000	2.8	3.6	3.1	3.9	0.3	2.0	-0.1	0.5	1.0	4.2	0.9	-1.2	0.7	4.9	2.0	1.2	2.6	1.1	1.7	2.2	3.7	0.0
UBD	N49629	0.000	-0.1	3.8	4.7	2.1	0.0	2.6	2.7	4.4	1.8	2.4	0.5	2.5	0.0	4.9	1.9	1.6	1.1	-0.2	-0.3	2.1	1.0	1.1
MMP7	AA031513	0.000	0.7	0.4	1.4	1.8	1.8	1.5	0.9	1.4	0.9	1.9	1.6	1.4	0.5	2.5	3.2	3.5	3.7	0.1	3.6	3.3	4.1	0.6
WNT2	H04382	0.000	NA	1.3	NA	4.5	2.8	0.3	NA	2.2	1.5	0.1	-0.1	2.5	0.9	0.2	NA	NA	NA	NA	1.9	2.7	4.2	0.5
IFITM1	AA058323	0.000	1.3	NA	1.8	1.7	0.3	3.9	2.4	2.1	2.2	1.2	1.1	0.7	0.6	2.3	1.1	1.9	1.4	3.2	1.4	2.1	2.3	0.2
CDH3	AA425556	0.000	2.0	0.4	1.8	3.9	2.0	2.6	0.3	0.8	1.2	1.3	2.3	1.7	1.1	0.1	NA	2.9	NA	0.9	NA	2.4	3.7	0.2
CDC2	AA598974	0.000	1.4	1.8	2.4	1.6	0.5	0.9	1.7	NA	2.3	2.3	1.2	0.7	2.7	2.5	1.8	2.6	NA	-0.1	0.6	2.0	1.8	1.4
WNT2	W93113	0.000	0.8	NA	-1.4	2.2	2.0	NA	0.6	1.6	2.6	-0.5	1.0	NA	NA	1.0	NA	NA	NA	0.6	3.0	4.4	4.6	NA
CXCL6	AI668847	0.000	3.6	NA	0.4	1.7	NA	-0.4	0.8	1.4	0.6	0.5	2.5	3.7	2.8	NA	1.2	NA	-1.2	-1.0	NA	3.6	1.5	3.6
IFITM2	AA985421	0.000	2.1	1.7	0.6	1.2	0.6	2.6	2.2	1.9	1.5	1.4	1.5	0.9	1.1	1.9	0.3	1.7	1.0	2.4	1.5	2.0	1.9	0.7
UBE2C	AA430504	0.000	1.2	1.0	1.3	2.7	1.4	2.2	2.5	2.1	2.4	0.9	0.8	1.6	1.2	0.7	2.6	2.5	1.1	0.3	0.9	1.0	1.4	1.0
SPARC	R52908	0.000	2.9	0.6	0.5	1.5	1.7	1.4	0.1	2.1	1.1	-0.2	2.6	1.3	-0.7	1.1	1.5	1.9	2.4	1.4	2.3	2.5	2.7	1.2

IFITM3	AA464417	0.000	2.0	1.5	0.6	1.1	0.4	2.8	2.6	1.9	1.6	1.5	1.5	0.9	0.8	1.8	0.1	1.4	0.8	2.5	1.2	2.1	1.8	0.7
IRAK3	AA885124	0.000	NA	NA	NA	NA	NA	-0.1	NA	NA	1.2	0.3	NA	NA	4.7	0.8	NA	NA	NA	0.1	NA	2.6	NA	1.8
WNT2	T99653	0.000	0.1	1.1	1.0	1.2	1.3	0.9	0.4	1.4	0.7	2.1	1.4	1.2	0.6	2.5	1.9	3.0	2.7	0.1	1.6	2.1	3.5	0.4
SLC2A1	AA679565	0.000	0.9	-0.7	NA	0.8	1.6	2.5	0.9	NA	1.6	0.2	1.6	-0.1	0.4	1.2	NA	4.7	4.5	NA	0.6	0.6	2.9	1.4
IFITM2	AA862371	0.000	1.9	1.7	0.9	0.9	0.6	2.8	2.1	1.5	1.4	1.2	1.2	0.7	0.9	1.6	0.1	1.6	1.5	2.4	1.5	1.7	1.4	0.8
ST00A9	AA864554	0.000	2.2	1.3	1.9	2.8	0.3	0.7	0.2	1.0	2.0	1.1	2.1	0.9	NA	1.2	0.4	0.6	NA	1.1	0.9	2.1	3.5	1.4
ETV4	H27201	0.000	-0.3	-0.3	0.5	3.9	2.9	1.0	0.3	0.6	0.4	0.3	0.2	3.0	-0.2	0.1	3.6	1.8	2.5	2.6	0.6	1.4	2.8	2.4
COL5A2	AA461456	0.000	2.5	1.3	0.8	0.7	0.7	NA	0.3	2.0	1.4	-0.4	3.2	0.7	0.9	0.6	2.2	1.3	1.5	2.1	2.0	2.1	1.9	0.9
CXCL10	AA878880	0.000	2.5	2.9	5.3	1.0	-2.5	3.1	0.4	2.0	3.6	1.9	-0.5	0.9	1.0	3.3	NA	NA	-1.7	0.1	-0.6	1.3	1.5	1.5
TCN1	AA155695	0.000	0.8	0.4	0.2	2.7	1.0	0.5	0.4	0.7	0.9	4.5	0.3	0.8	-0.6	3.7	NA	NA	NA	0.2	0.5	0.5	4.6	3.6
CA8	W93386	0.000	0.5	-0.2	NA	NA	2.3	-0.9	4.6	-1.8	5.0	0.4	1.9	3.1	0.2	NA	NA	NA	NA	NA	NA	NA	0.5	1.6
DDIT4	AA447746	0.000	0.8	0.0	1.4	0.4	2.1	2.4	0.0	0.7	1.1	2.0	1.5	0.6	0.7	1.5	1.4	2.1	2.0	1.1	0.9	2.1	2.4	
CCL20	AI285199	0.000	0.6	2.5	3.1	2.5	1.5	3.4	2.4	1.6	0.3	0.4	1.0	-0.4	0.2	2.3	2.4	-0.1	1.3	-1.9	1.6	3.7	1.0	-0.3
RRM2	AA187351	0.000	1.9	1.5	1.7	1.9	0.2	1.4	0.6	1.9	2.0	0.9	0.3	0.2	2.3	0.6	3.8	2.2	1.7	0.0	1.3	0.3	0.5	1.4
AKT1	W77811	0.000	0.1	0.7	NA	0.9	0.9	1.3	0.7	0.9	0.9	2.2	1.6	1.2	0.5	2.4	NA	2.4	2.2	0.6	2.0	2.0	2.2	0.2
AGT	AI359985	0.000	0.3	0.7	1.4	2.2	1.2	2.9	0.5	1.3	0.3	0.8	0.2	1.4	1.4	0.2	0.3	3.0	1.9	1.3	1.4	2.2	2.4	1.1
SLC2A1	H58873	0.000	1.0	0.1	0.0	0.5	1.5	3.3	0.9	0.9	1.3	0.0	1.6	0.2	NA	1.3	3.7	3.9	2.0	-0.1	1.1	0.4	2.2	0.9
IFITM1	AA419251	0.000	1.0	0.6	1.6	1.2	-0.1	3.1	1.3	1.1	2.0	1.5	0.6	0.3	1.0	1.1	2.4	NA	0.4	2.7	1.0	2.2	1.7	0.1
PTTG3	AA430032	0.000	0.8	1.4	1.0	1.9	0.5	1.5	0.9	0.9	2.3	1.7	0.7	1.3	2.1	0.7	NA	1.9	1.7	0.4	0.9	1.7	1.1	1.1
ENC1	H72122	0.000	1.3	0.8	1.9	0.8	1.1	0.5	1.3	0.1	0.6	1.2	2.0	1.7	0.5	0.9	1.2	1.8	2.5	0.6	2.1	2.6	0.6	1.5
PPBP	W92812	0.000	1.0	0.0	1.4	0.5	0.2	1.0	1.0	0.4	0.9	0.3	1.3	0.1	0.8	1.3	NA	4.3	-0.6	4.1	0.2	6.0	0.5	
RASGRF2	R55406	0.000	0.7	NA	1.3	1.9	0.9	0.6	0.9	1.4	0.9	0.1	0.7	1.3	NA	0.7	0.8	0.4	4.9	1.1	0.6	1.9	1.7	1.6
MAD2L1	AA481076	0.000	1.9	2.0	1.8	NA	0.7	NA	0.8	NA	1.7	0.9	1.3	0.9	2.3	1.3	NA	1.9	NA	-1.0	-0.5	1.3	2.0	1.4
CDK4	AA486312	0.000	0.6	0.8	1.9	2.0	2.2	1.1	1.0	-0.3	1.4	2.4	0.7	1.3	1.5	1.0	1.5	1.0	0.8	0.9	0.9	1.6	1.1	0.7
CKS2	AA010065	0.000	0.6	1.1	0.6	2.3	0.8	1.1	2.3	0.8	2.2	0.9	0.8	0.7	1.9	0.7	1.3	1.2	0.2	0.6	1.4	1.7	1.9	1.2
ANLN	R16712	0.000	1.3	1.0	1.1	0.7	-0.2	0.5	1.7	2.3	1.7	1.5	0.8	0.0	2.1	1.9	2.1	2.4	3.1	-1.2	1.3	0.6	0.4	0.8
SFRP4	AA486838	0.000	0.0	0.0	3.7	1.2	1.3	0.5	0.5	0.3	1.7	NA	1.4	1.6	-0.5	0.4	NA	2.7	NA	0.9	1.5	2.2	2.7	0.4
IGFBP3	AA598601	0.000	1.8	1.3	1.1	1.0	2.1	0.8	1.1	0.9	1.9	0.8	1.6	1.9	0.0	1.4	1.3	-0.5	1.0	0.3	1.3	1.8	0.9	2.0
MMP10	AA857496	0.000	0.2	NA	NA	NA	1.3	-0.3	0.7	0.0	3.7	0.9	1.3	NA	0.2	0.5	NA	0.7	NA	NA	2.4	2.5	NA	2.4
PANX1	AA428957	0.000	0.8	-0.2	1.4	2.6	1.1	NA	0.0	NA	1.9	1.1	1.0	1.3	0.3	-1.5	NA	2.1	1.1	1.3	NA	1.7	3.4	1.7
CXCL2	R50407	0.000	0.8	2.2	3.5	-0.2	0.1	2.1	1.4	NA	1.9	2.0	0.8	0.2	0.2	1.9	0.5	2.5	NA	-0.1	-0.1	2.1	0.3	1.0
CAPG	AA486942	0.000	0.7	1.0	1.6	1.6	0.3	1.8	0.6	1.8	1.9	1.3	0.9	0.3	1.9	0.7	2.1	1.5	1.1	-0.2	0.6	0.9	1.5	1.5
COL11A1	BX109171	0.000	1.1	1.0	NA	NA	1.3	0.8	-0.1	0.3	2.4	0.2	3.2	1.6	0.0	-0.3	2.4	0.6	1.3	1.9	1.4	NA	2.4	0.3
ST00A8	AA086471	0.000	0.8	NA	NA	0.5	-0.2	NA	-0.2	1.0	1.9	0.4	2.4	0.4	NA	2.8	1.5	NA	3.5	1.0	-0.3	0.9	1.6	1.4
HSPC150	AA460432	0.000	0.0	0.6	1.2	2.1	0.4	0.7	1.5	1.6	2.3	1.1	1.1	0.6	1.2	0.2	0.1	1.3	0.5	0.0	0.5	1.7	1.4	5.7
BMP4	AA463225	0.000	0.5	0.5	0.7	2.5	1.7	1.4	0.0	0.1	0.9	1.4	1.8	0.2	2.4	0.7	NA	1.1	2.8	0.9	1.1	0.4	2.2	0.2
C6orf173	W90323	0.000	1.1	1.0	1.7	2.9	0.4	1.1	0.0	0.4	1.0	1.6	0.6	0.8	1.6	0.2	2.1	NA	1.7	0.3	0.5	0.3	2.8	1.6
LYZ	N63943	0.000	-0.4	0.8	1.4	0.2	-1.2	-0.1	-0.7	1.1	0.9	3.1	0.8	-0.7	2.9	3.9	2.5	0.9	1.4	0.5	2.4	0.3	2.6	2.2
DCC1	AA843451	0.000	0.3	1.3	0.0	2.0	1.5	1.0	1.2	1.0	1.4	NA	0.3	0.8	1.5	1.8	NA	0.8	NA	0.5	NA	1.4	2.2	1.3
SNAP25	AA663884	0.000	1.1	1.7	1.8	2.0	1.7	1.3	1.3	0.8	0.9	1.3	0.3	0.7	0.8	0.6	1.7	1.0	0.8	0.9	0.8	1.1	1.5	0.6
C10orf3	AA131909	0.000	0.4	1.3	1.9	1.3	0.2	0.5	1.2	1.0	2.1	0.3	0.5	0.5	1.6	0.9	NA	2.8	2.1	0.6	0.4	1.2	1.4	1.3

FZD3	H84795	0.000	0.2	0.0	2.1	0.8	2.3	0.5	0.9	2.7	1.7	0.5	0.2	1.3	0.5	NA	0.7	NA	1.3	0.7	NA	3.7	0.1
CKS2	AA397813	0.000	0.7	0.9	1.0	2.2	0.8	1.1	1.8	0.6	2.1	1.2	0.8	0.5	1.4	0.4	1.5	0.6	0.9	0.5	1.3	1.6	1.7
RAB31	T96083	0.000	2.4	NA	NA	1.0	0.2	2.4	-0.4	NA	0.9	1.0	2.2	-0.1	-0.4	1.4	2.0	0.8	2.5	0.2	1.1	1.2	2.2
MMP1	AA143201	0.000	1.8	1.1	1.0	1.6	0.3	0.8	0.4	1.4	1.6	1.7	0.7	0.5	0.9	1.1	1.0	1.5	1.2	0.2	1.2	0.7	2.1
PTTG3	AI362866	0.000	0.4	1.3	1.3	1.0	0.7	1.5	0.2	0.1	2.1	NA	0.8	1.3	1.7	0.2	0.9	3.9	NA	0.2	1.1	1.2	1.1
MLF1IP	AA431741	0.000	0.8	0.3	1.3	NA	-0.4	0.1	0.6	3.2	1.9	1.0	0.0	0.5	0.6	0.5	1.1	NA	NA	-0.5	4.2	0.7	3.4
ATP6V1F	AA664077	0.000	1.2	1.2	0.8	1.3	1.3	1.5	0.3	0.6	1.1	1.7	0.8	1.0	1.7	0.6	2.3	1.4	-0.1	1.4	0.4	0.8	1.4
FN1	R62612	0.000	2.5	0.2	0.2	1.1	1.8	-0.1	2.7	2.8	2.1	-0.1	4.5	-0.5	-1.3	0.5	2.2	-0.8	-0.9	-0.1	2.7	1.4	1.9
CCNA2	AA608568	0.000	1.3	1.1	1.8	0.7	0.0	0.7	1.0	2.2	1.4	1.5	0.9	0.2	2.5	1.0	NA	2.7	0.8	NA	0.2	0.4	0.7
BUB1	AA446462	0.000	0.8	1.6	2.0	NA	0.0	-0.3	1.5	3.2	3.2	1.3	0.2	0.2	0.9	0.7	NA	1.7	0.9	-0.4	1.6	0.6	1.0
NA	AA668189	0.000	0.6	0.5	0.6	1.2	1.2	1.2	1.1	1.6	1.5	1.9	1.0	1.1	0.9	-0.2	1.6	1.5	0.9	0.4	0.8	1.4	1.5
KIAA0101	W68220	0.000	-0.1	0.3	NA	NA	0.7	0.6	1.3	3.6	2.1	0.8	0.5	0.4	1.5	0.9	NA	NA	-0.3	1.2	1.8	NA	1.8
APOC1	AA873159	0.000	-1.0	1.5	0.9	0.3	-0.9	0.8	0.1	1.3	1.3	-0.1	0.8	0.8	0.6	2.0	4.6	-0.8	4.2	NA	2.0	0.4	2.0
SERPINH1	R71440	0.000	1.1	NA	0.2	1.0	1.0	1.0	1.0	0.9	0.6	0.3	1.5	0.8	0.7	0.9	0.8	1.8	3.4	0.4	1.3	1.6	1.7
LAMP2	N77754	0.000	0.0	0.4	0.4	0.4	0.8	1.0	1.3	2.1	0.9	0.8	0.6	1.0	0.6	1.4	5.9	NA	2.1	0.3	0.3	0.6	0.9
ALDH1A3	AA455235	0.000	0.3	0.0	0.4	2.7	2.2	0.9	0.4	2.6	2.9	0.7	0.2	0.9	1.9	0.6	-0.5	1.8	0.8	-1.5	1.6	2.3	2.0
MSH3	AA421716	0.000	2.3	0.7	1.8	3.6	0.7	NA	-0.4	0.8	0.8	2.9	0.1	0.0	0.2	NA	0.5	1.7	0.7	NA	0.2	0.8	2.0
NOLA2	AA485675	0.000	0.3	0.6	0.6	1.9	1.7	1.2	1.2	0.5	0.9	0.8	0.7	1.3	1.3	0.5	0.8	0.5	1.2	1.2	1.7	1.8	1.3
CGA	AA677403	0.687	-0.4	-0.8	0.8	1.3	1.0	0.2	1.8	11.8	0.1	0.5	0.6	0.4	0.8	0.3	NA	0.9	0.9	0.7	-0.3	0.6	0.3
G1P2	AA406020	0.000	0.5	0.0	2.3	0.7	-1.7	0.1	0.5	0.5	2.5	0.6	0.6	-1.2	1.1	2.6	2.8	3.6	1.3	1.2	0.8	0.8	1.9
GSTP1	R33755	0.000	0.9	-0.2	0.2	1.7	-1.3	1.5	-0.1	1.1	0.9	1.6	0.7	1.4	1.5	0.2	1.6	1.9	2.2	1.0	1.8	1.4	1.4
MCAM	AA497002	0.000	1.1	0.3	NA	1.2	1.3	0.8	1.9	2.2	1.0	0.2	3.3	-0.5	NA	0.2	0.6	0.0	2.3	1.1	1.8	0.6	0.9
RHEB	AA482117	0.000	1.1	1.0	0.7	0.7	0.1	1.2	1.6	2.1	0.9	0.8	1.3	0.1	1.8	1.8	1.6	1.2	0.9	0.7	0.5	0.8	1.2
MMP14	N33214	0.000	0.3	0.5	1.6	1.2	0.5	0.5	0.4	0.7	1.5	-0.4	2.1	0.6	0.6	0.0	3.0	1.1	NA	NA	2.5	0.6	1.3
CEBPB	H26183	0.000	0.6	0.1	0.5	1.6	1.6	1.5	1.2	1.1	1.6	0.9	0.6	1.6	0.1	1.3	0.3	0.9	1.0	1.5	1.5	0.7	1.3
CDH11	H96738	0.000	0.6	1.9	NA	0.7	0.8	0.8	0.1	1.5	0.3	0.3	1.9	0.6	-0.5	1.1	2.1	1.0	3.4	0.6	1.9	1.2	1.2
RBBP7	AA130591	0.000	0.4	0.0	1.9	1.5	1.4	1.5	0.8	1.9	0.7	0.3	0.5	1.8	0.2	1.3	0.2	3.9	0.1	0.7	0.3	1.2	0.9
ARPC1B	AA188155	0.000	0.8	0.6	0.6	1.1	0.9	0.7	1.2	0.4	2.0	0.6	1.1	1.5	1.2	1.4	1.3	0.4	0.3	0.7	1.7	1.4	1.1
TDGF1	AA101473	0.000	0.2	-1.4	0.8	1.8	2.5	0.9	1.1	0.9	0.5	1.7	1.2	2.8	0.9	0.8	2.6	1.4	NA	0.3	0.2	0.1	1.0
IL1A	AA936768	0.000	0.2	0.6	NA	NA	NA	0.6	0.3	0.9	3.4	0.3	0.1	2.2	0.8	1.1	NA	NA	NA	2.1	0.4	NA	NA
NA	AA707550	0.000	0.4	2.3	0.7	1.0	0.9	0.9	1.1	0.7	1.0	1.8	0.8	0.8	1.0	0.0	1.2	1.5	1.3	0.7	1.1	0.4	0.9
IER3	AA480815	0.000	0.9	1.0	0.4	0.0	1.0	1.7	1.7	0.8	1.5	0.5	1.3	0.5	0.1	1.0	NA	2.3	1.0	1.8	0.9	1.4	0.7
ANXA3	H67075	0.000	1.1	0.9	1.6	0.3	1.5	0.8	0.4	0.5	1.1	1.8	0.7	0.0	0.8	0.8	1.4	1.5	2.2	0.5	1.5	-0.1	2.0
CCNB2	AA774665	0.000	1.1	0.4	1.0	1.4	0.1	1.6	1.0	1.6	1.9	1.0	0.5	0.7	1.5	0.4	0.9	2.1	0.9	0.2	0.3	1.0	1.3
USP24	W86608	0.000	0.0	0.0	0.6	2.0	1.1	0.4	0.7	0.7	0.3	0.4	1.1	2.2	0.3	0.1	1.4	3.1	1.3	0.1	1.4	2.3	0.9
PROX1	H82137	0.000	1.6	NA	NA	NA	NA	3.4	NA	0.1	NA	-0.4	0.6	2.7	NA	0.4	NA	NA	NA	NA	0.6	-0.7	2.0
HNRPLL	H95141	0.000	-0.2	1.2	0.6	1.6	1.4	1.2	1.5	1.0	1.0	1.0	0.7	1.4	0.9	0.6	0.9	0.5	1.1	1.0	1.1	1.2	0.8
WNT10B	AI652223	0.000	-0.2	NA	0.3	NA	0.8	0.3	0.1	0.8	2.0	-0.5	-0.5	1.4	1.3	2.5	NA	NA	NA	0.2	0.6	2.6	2.6
COL1A2	AA490172	0.000	2.6	0.2	0.1	1.1	1.1	-0.1	-0.1	1.7	0.8	-1.8	2.4	0.5	-0.9	0.2	2.0	1.3	2.8	1.5	1.6	1.2	3.0
SLC2A1	R17667	0.000	1.1	-0.4	0.1	1.3	1.4	1.2	0.4	0.7	0.8	0.2	0.8	0.3	0.4	1.0	2.3	3.0	1.3	0.7	1.2	-0.1	2.8
CAB39L	T55806	0.000	1.0	1.1	0.7	0.9	0.7	2.2	0.3	1.2	1.2	2.5	1.0	0.7	1.5	-0.1	0.6	1.7	0.6	0.7	0.8	0.6	0.5

MRPS30	N33236	0.000	-0.1	1.6	1.1	1.2	0.8	0.0	0.9	1.2	4.3	0.5	0.2	0.8	0.9	0.3	3.0	0.5	NA	-0.2	0.6	0.9	1.0	0.7
MYC	W87741	0.000	0.2	1.1	1.0	1.2	1.2	1.4	1.5	1.3	0.3	1.9	0.9	1.6	0.8	1.3	0.3	NA	NA	0.4	0.4	2.2	-0.4	0.5
HIG2	AA670144	0.000	0.7	0.1	1.3	2.2	2.4	0.8	1.2	1.4	1.3	0.1	1.1	2.2	0.2	0.1	0.8	NA	-0.2	0.3	-0.8	1.2	2.3	1.3
RAB32		0.000	0.0	-0.1	NA	1.6	2.0	0.6	1.4	0.7	1.6	0.2	0.0	1.2	0.2	1.2	3.6	0.1	NA	0.7	0.5	2.1	0.0	1.5
MCTS1	R70784	0.000	0.6	0.5	0.5	1.4	1.7	0.6	1.0	1.7	0.9	1.3	0.5	0.8	1.8	0.0	1.4	1.3	1.2	0.3	0.7	0.5	1.2	0.8
BMP7	W73473	0.000	1.0	0.3	NA	3.4	0.3	0.8	0.4	-0.2	0.4	0.4	0.7	0.5	0.5	0.9	1.7	3.4	NA	0.8	0.8	1.5	NA	0.2
APCDD1	W76603	0.000	0.3	0.5	1.2	3.4	2.5	1.3	2.2	0.7	0.5	1.2	-0.1	1.4	1.6	0.0	0.8	0.2	0.9	0.6	0.5	0.5	-0.3	0.8
RIS1	AA127069	0.000	0.3	0.2	2.0	2.3	1.1	0.8	1.2	0.9	1.0	0.4	0.6	0.2	-0.4	0.3	0.3	1.4	2.1	0.2	1.6	1.9	1.7	0.5
RAB31	AA449333	0.000	1.0	0.3	0.4	0.8	0.0	0.5	-0.2	1.9	0.9	0.1	2.1	-0.2	-0.5	1.3	1.8	0.6	2.6	NA	1.9	1.5	2.5	0.7
MFAP2	N67487	0.000	-0.1	0.2	0.0	3.8	1.5	0.4	-0.2	0.4	2.5	0.1	0.6	1.1	0.4	0.2	0.8	1.4	2.0	0.2	1.4	0.3	2.7	0.4
MRPL50	AA700090	0.000	1.9	1.4	1.6	1.3	0.8	1.2	1.1	0.2	1.8	1.3	0.4	0.5	2.4	0.9	0.5	-0.7	0.4	-0.5	0.9	0.9	0.8	1.2
BHLHB2	AA026120	0.000	1.1	1.6	1.8	0.6	1.0	1.2	0.8	0.7	0.9	0.2	2.5	0.4	0.1	0.6	0.9	1.1	0.6	0.9	1.6	0.8	0.9	0.1
COL6A3	R62603	0.000	1.2	-0.4	0.3	0.8	0.2	1.4	-0.2	2.5	0.8	-1.3	1.5	0.5	-0.8	0.6	1.0	1.4	2.2	0.6	1.5	1.8	4.1	0.4
CDC25A	AA913262	0.034	0.5	NA	0.0	1.6	0.4	5.9	0.3	1.0	0.7	0.4	0.1	0.3	3.0	0.6	0.4	0.7	0.3	0.0	NA	0.8	0.6	0.5
TPD52L1	AI014441	0.000	1.3	0.1	0.5	4.6	1.8	0.8	0.8	1.3	1.0	0.6	0.6	1.4	0.6	-0.1	0.6	0.1	0.6	-0.1	1.0	1.0	NA	0.7
ANGPT1	AA679746	0.000	0.0	-0.4	NA	NA	3.7	NA	0.7	NA	NA	-0.6	-1.2	3.7	NA	0.9	NA	NA	NA	1.0	NA	NA	NA	1.2
CCL18	AA495985	0.000	0.3	-0.1	3.3	NA	-0.2	1.6	-1.0	1.1	3.4	-0.1	0.9	0.3	-0.6	0.8	2.9	-0.6	1.0	0.0	1.6	0.1	2.3	1.8
PCSK9	W92036	0.000	1.6	0.7	1.7	1.1	1.0	0.1	0.7	-0.3	0.6	0.7	0.9	-0.2	0.2	0.2	0.3	3.9	1.9	1.3	0.7	0.2	0.9	1.7
NA	R83270	0.000	0.2	NA	0.5	1.3	1.3	1.1	1.8	0.4	1.1	1.6	1.2	0.8	NA	0.4	1.6	1.8	0.4	0.7	0.7	1.3	-0.5	0.5
CASP7	T50828	0.000	0.7	0.0	0.1	1.8	0.7	0.3	0.1	1.2	0.3	0.3	1.0	0.6	0.0	0.0	1.2	2.2	2.6	NA	1.9	1.6	2.0	0.4
PCDHA6	AI968755	0.000	1.6	NA	0.2	1.9	1.9	4.6	-0.2	-0.1	0.5	-0.3	0.1	1.6	-0.3	-0.3	NA	NA	NA	-0.2	2.8	1.2	NA	0.4
SCD	R00707	0.000	0.3	0.4	2.9	1.4	1.1	0.1	0.9	1.7	1.0	-0.1	0.5	0.0	-0.2	-0.1	1.9	3.3	2.4	0.5	0.8	-0.6	0.2	1.7
NMIE1	AA644092	0.000	0.2	0.5	1.6	2.3	2.1	0.7	0.3	0.7	1.1	0.9	0.4	0.7	0.1	0.2	0.3	0.9	1.8	0.3	0.5	1.6	1.4	1.3
BIRC4	AI628066	0.000	0.3	NA	1.0	NA	4.5	0.3	0.1	2.6	1.2	0.1	0.3	NA	1.6	-0.1	0.7	0.4	0.2	0.0	1.1	0.5	NA	1.6
CDC6	H59203	0.000	0.8	0.5	0.6	0.9	0.5	1.2	1.2	1.1	0.5	1.2	0.3	0.2	2.4	0.4	1.1	1.2	1.5	0.5	0.1	0.2	3.1	0.5
DXS9879E	AA480035	0.000	0.3	0.0	0.5	1.7	1.7	0.9	1.5	0.4	1.0	1.1	0.7	1.6	0.6	0.6	0.7	1.6	1.2	0.3	0.8	1.4	0.3	0.8
OVGP1	AA863449	0.000	1.2	1.4	-0.1	0.5	0.7	2.9	2.3	1.5	0.0	0.9	0.8	0.9	2.2	0.1	2.3	-0.5	1.2	-0.1	0.9	0.3	0.6	-0.3
STK6	R19158	0.000	0.2	0.1	1.8	1.8	0.2	1.4	1.3	1.7	1.6	0.3	0.1	0.8	0.4	0.1	1.9	2.7	0.9	-0.2	0.8	0.8	0.0	1.0
CSPG2	AA101875	0.000	1.3	0.6	0.3	-0.3	1.4	0.7	0.1	2.2	1.0	-0.1	1.7	0.9	-0.3	0.2	2.0	1.3	1.3	1.1	0.9	1.6	0.9	0.6
GAD1	AA018457	0.000	-0.1	-0.9	NA	NA	1.6	-0.4	0.4	1.6	2.6	0.7	0.4	2.4	NA	0.1	NA	NA	NA	-0.4	1.7	2.6	NA	NA
CNTN6	N30615	0.031	1.2	-0.1	-0.3	NA	1.1	NA	NA	NA	-0.4	0.8	-0.5	3.9	-1.2	0.3	3.9	NA	1.5	0.0	3.4	1.0	-0.5	
CYP1B1	AA448157	0.033	-0.5	-0.1	0.4	0.8	0.4	NA	0.1	0.8	1.6	0.0	0.6	0.4	NA	0.9	3.3	NA	0.8	NA	-0.6	5.0	2.2	-0.2
OAS3	AI357590	0.000	0.6	0.7	1.3	0.8	-2.4	1.6	0.4	2.3	1.7	0.0	-0.2	-0.9	0.3	1.6	2.2	2.8	2.4	0.2	1.8	-0.1	1.3	0.9
ECM1	N79484	0.000	0.7	0.9	1.6	0.5	0.3	1.2	0.0	0.2	0.2	0.5	1.3	0.8	-0.4	1.0	2.7	2.2	2.4	0.4	1.0	-0.5	1.6	0.8
TFF1	AA573793	0.000	0.8	1.2	1.0	0.3	1.0	0.5	1.1	0.2	0.7	0.3	2.2	-1.0	3.2	0.2	1.3	1.4	0.0	0.3	0.6	-1.6	1.3	4.3
PSMB5	AA864479	0.000	1.7	1.9	3.3	0.0	1.2	0.0	0.8	1.1	-0.2	2.9	1.6	0.8	1.2	-0.4	-0.3	1.8	1.4	-0.4	0.7	0.6	-0.8	0.4
TNFAIP6	W93163	0.000	2.4	0.2	-1.3	2.8	0.6	0.2	0.3	2.0	2.3	0.3	0.2	1.4	0.6	1.1	NA	NA	NA	-2.5	0.9	1.1	2.8	1.2
DKC1	AA052960	0.000	0.8	0.3	NA	0.9	1.3	1.2	1.6	1.5	0.6	0.3	0.2	0.9	0.7	0.4	1.9	2.0	0.4	0.5	0.4	1.1	0.8	0.8
CA9	AI023541	0.000	0.3	1.6	2.9	0.4	0.9	0.7	0.4	0.0	0.1	0.9	1.3	1.3	2.0	0.5	0.3	1.4	0.4	0.7	0.1	1.1	1.5	0.1
FAIM	AA099588	0.000	0.0	0.8	1.4	1.6	1.0	-0.8	0.8	NA	1.6	0.7	0.0	0.3	1.1	NA	2.8	0.7	NA	-0.4	1.3	1.2	1.0	1.0
TNC	T77595	0.000	3.7	0.2	1.1	1.1	-0.8	0.2	-0.1	1.9	2.3	0.0	0.9	-1.3	0.1	0.5	-0.8	NA	1.5	0.8	0.6	1.7	3.5	0.7

C9orf140	AA088458	0.000	0.7	1.2	0.6	2.3	0.4	1.7	1.2	0.0	0.9	1.1	0.7	0.7	1.0	0.2	1.3	0.8	0.1	0.3	0.1	1.5	1.0	1.1
CCND1	AA487700	0.000	0.4	0.1	0.6	1.1	1.0	1.4	0.6	1.1	0.8	0.4	0.8	0.3	1.7	0.7	1.8	1.4	1.4	0.6	0.3	0.5	0.8	0.9
CXorf35	AI079530	0.000	0.5	0.1	0.3	2.7	0.7	1.1	0.7	1.0	1.0	0.9	0.9	0.4	1.1	0.1	0.7	1.4	0.8	0.6	0.7	1.2	1.0	1.1
UNC84A	AA055971	0.126	NA	NA	1.2	0.6	0.1	NA	NA	NA	0.2	NA	0.6	0.2	5.8	2.3	1.7	0.7	NA	0.3	-0.5	-0.2	-0.2	0.1
DCTD	AA448207	0.000	0.3	1.0	0.9	0.8	0.9	0.8	0.8	0.7	0.8	0.7	0.1	1.0	1.8	0.8	0.7	2.1	0.8	0.5	0.6	1.0	0.8	1.1
SLC2A8	T59518	0.000	0.3	0.6	NA	1.5	0.9	2.7	0.9	0.8	0.4	0.1	0.4	0.5	NA	0.3	0.8	2.8	1.7	0.7	0.2	0.9	0.4	0.5
CCT4	AA088226	0.000	0.3	1.7	1.1	0.9	0.4	0.5	1.4	1.1	1.4	1.5	0.5	0.5	1.6	0.3	0.7	0.8	0.9	0.7	0.5	0.6	0.9	0.5
EPHB3	AA455591	0.000	-0.2	0.6	1.1	1.3	2.6	0.1	1.2	0.7	0.5	0.9	1.0	2.5	1.0	1.2	1.4	-0.1	0.3	1.5	0.1	1.3	-0.8	0.5
RGST19IP1	AA133577	0.000	0.4	1.0	0.8	1.7	0.9	0.4	1.0	1.5	1.4	1.3	0.5	0.4	1.4	0.2	0.6	1.1	1.2	0.0	0.6	0.5	1.1	0.7
CXCL5	BX108729	0.000	1.0	0.4	0.7	0.4	NA	0.2	0.5	0.2	0.8	0.2	1.1	0.3	2.8	0.7	0.9	2.3	1.7	0.0	1.0	0.0	0.4	2.2
DLX4	AA256304	0.000	1.8	NA	NA	1.4	0.6	1.2	0.2	-0.3	0.3	0.9	0.5	0.4	NA	0.7	4.8	NA	NA	0.0	0.6	0.5	0.1	0.8
TIMM10	AA670296	0.000	0.7	NA	1.3	1.2	1.3	0.5	0.5	0.8	0.7	0.9	0.0	0.6	1.9	0.1	1.4	1.1	2.5	0.0	0.7	-0.1	0.6	1.1
CDK7	R22625	0.000	0.8	0.5	0.8	0.3	0.8	0.1	1.0	1.1	0.7	1.2	1.3	0.9	1.5	1.2	1.1	0.9	1.1	0.4	0.7	0.7	0.7	0.7
STK6	R19158	0.000	0.3	0.5	0.7	1.2	0.4	0.3	0.9	1.0	1.1	0.8	0.7	0.7	1.0	0.2	1.9	2.2	1.5	0.1	1.3	0.6	0.5	0.7
BAAT	R98624	0.758	0.0	0.0	NA	1.8	-0.1	-0.1	0.1	-0.3	0.1	-0.1	0.1	0.3	0.4	0.4	4.3	NA	7.5	-0.6	1.2	1.6	0.1	0.1
TOP2A	AA504348	0.000	0.8	1.0	0.9	0.3	0.4	0.7	1.4	1.4	1.7	1.0	0.5	0.5	1.3	0.6	1.7	2.0	0.5	-0.7	0.3	0.6	0.8	0.5
AREG	AA857163	0.000	0.4	1.0	1.0	0.7	0.6	2.1	2.0	2.6	0.7	0.0	-0.9	1.7	1.3	0.7	0.1	0.6	1.9	0.0	1.4	1.2	-0.8	0.0
HSPH1	AA485036	0.000	-0.8	0.9	1.3	1.2	0.4	1.0	2.8	1.8	0.8	-0.3	0.5	0.4	0.5	-0.4	1.8	1.2	2.0	0.6	1.4	0.3	1.6	-0.4
TRIM32	AA055503	0.000	0.7	0.8	0.5	1.6	0.9	2.5	0.7	0.7	0.6	0.3	0.7	0.2	0.6	-0.4	2.1	1.2	NA	0.1	0.9	0.7	1.1	1.1
CCT6A	AA872690	0.000	0.7	NA	0.8	0.6	0.3	0.7	1.4	1.9	0.4	1.0	1.0	0.4	1.3	1.1	0.7	0.7	1.3	0.3	-0.3	1.0	1.2	1.2
LYN	AA918079	0.000	0.9	1.1	0.3	0.5	-0.1	0.8	0.5	1.4	0.6	0.9	0.7	0.2	1.8	1.1	0.9	-0.1	2.3	-0.4	1.1	0.9	1.6	1.0
MGC9850	AA136527	0.000	1.0	0.7	0.2	1.4	1.6	0.8	1.0	1.4	0.5	1.3	1.0	0.8	1.2	0.6	1.6	0.2	0.0	0.3	1.0	0.4	0.7	0.7
NA	R23055	0.000	0.2	0.3	1.7	2.9	0.5	0.2	0.8	0.2	1.0	0.0	-0.2	1.2	-0.1	0.3	3.0	1.4	NA	0.3	0.2	0.7	1.2	1.7
ZNF281	N47468	0.000	0.6	-0.1	0.8	NA	0.4	-0.2	0.7	1.3	0.6	0.2	1.6	0.2	-0.2	1.1	1.8	-0.2	2.7	1.0	1.8	1.8	1.5	-0.1
MID1IP1	R91137	0.000	0.5	0.6	0.0	1.7	1.7	1.0	0.9	0.2	0.6	0.3	0.4	1.1	0.9	0.6	-0.1	1.9	1.5	0.6	0.7	1.3	1.2	0.8
CCND1	R81200	0.000	0.2	0.3	0.9	1.1	1.4	0.9	0.7	1.6	0.6	0.8	0.8	0.4	1.0	1.4	0.9	1.7	1.4	0.7	1.6	-0.5	-0.6	0.9
MGC2714	AA186873	0.000	0.6	0.6	1.3	1.2	1.0	0.2	-0.1	0.7	1.5	1.1	0.4	0.8	1.5	-0.4	1.3	2.4	2.0	-0.2	0.2	0.2	0.7	1.1
CBFB	AA187148	0.000	1.1	0.8	1.3	0.4	0.2	0.8	1.1	0.5	0.8	1.1	0.8	0.7	0.8	1.0	1.5	1.2	1.6	0.2	0.4	0.4	0.7	0.5
COL15A1	AA464342	0.000	2.4	0.2	0.0	1.9	0.0	0.7	0.2	1.9	0.9	-0.5	1.4	0.9	0.3	0.4	0.4	0.5	1.2	-0.4	0.7	1.9	2.1	0.8
FKBP2	R75819	0.000	2.0	0.3	0.1	0.5	NA	-0.3	2.0	1.1	1.1	0.1	3.5	-0.6	-0.3	-0.1	1.7	-0.3	2.0	-0.1	1.9	0.8	1.2	0.6
LATS2	N64139	0.000	NA	-0.6	NA	2.0	0.3	NA	0.7	0.9	0.7	1.0	0.6	0.5	NA	0.4	1.3	0.6	0.4	NA	1.3	1.1	1.7	0.9
NA	AA448396	0.000	0.0	0.9	0.4	1.1	0.8	1.1	1.9	1.1	1.3	0.6	0.5	0.7	1.6	-0.1	0.7	0.7	0.5	0.4	0.9	1.2	1.0	0.6
MMP9	AA425227	0.000	1.8	NA	0.7	2.1	0.7	1.1	0.5	1.4	0.0	1.5	1.7	1.1	-0.2	1.2	NA	NA	-2.3	0.1	-0.1	0.9	1.7	1.4
SYNPR	N35684	0.081	0.6	0.4	-0.3	2.9	2.5	-1.1	0.2	NA	0.0	-0.7	0.9	1.4	0.0	-0.1	NA	NA	-0.2	NA	4.7	1.8	0.6	0.6
RAN	AA456636	0.000	0.5	0.6	0.7	0.6	0.5	0.8	1.4	1.4	0.5	1.7	0.8	0.7	1.3	1.0	1.0	0.7	0.7	0.2	0.2	1.0	0.8	0.7
NA	BX113955	0.000	0.2	2.2	NA	-0.5	0.9	-0.5	1.0	1.4	0.1	0.8	0.1	0.3	1.1	2.3	1.3	1.4	NA	NA	NA	1.1	0.5	0.7
NUTF2	N75595	0.000	0.3	0.6	0.3	0.9	1.0	0.9	1.2	1.0	1.5	0.4	0.2	0.9	0.9	0.7	0.8	1.2	1.0	0.6	0.0	0.8	1.0	1.1
VNN1	AA983530	0.126	0.8	0.6	0.5	0.1	0.2	0.5	-0.4	1.0	0.0	0.4	0.4	-0.6	0.9	1.4	1.2	6.5	NA	0.4	NA	1.1	0.8	0.0
PRKD1	N53380	0.000	-0.2	-0.1	0.7	NA	3.4	0.9	0.0	0.3	1.7	0.3	-0.1	1.1	-0.2	1.3	NA	-0.5	NA	-0.2	-0.1	2.0	2.9	1.7
MEOX2	H25223	0.000	0.1	0.2	-1.0	NA	1.4	NA	0.4	2.1	NA	0.1	1.6	2.1	0.1	NA	NA	NA	NA	-0.9	1.0	1.0	NA	2.9
ZA20D2	AA463929	0.423	1.2	0.2	7.1	0.6	0.1	0.2	0.1	0.9	0.1	0.5	0.7	0.3	NA	0.2	NA	1.4	NA	NA	0.8	-0.2	-0.2	0.2

G1P3	AA432030	0.000	1.2	0.9	0.9	1.0	-0.8	0.6	0.0	0.9	2.1	1.1	0.2	-0.4	0.2	1.2	1.9	1.1	1.3	0.5	0.9	0.2	1.6	0.7
LUM	AA453712	0.000	0.9	0.7	0.6	0.2	1.1	-0.2	0.0	2.6	0.5	0.3	2.1	0.4	-0.4	0.7	0.5	0.1	0.8	0.8	1.5	2.0	2.4	0.0
CDKN1C	A1676118	0.000	0.1	1.1	1.0	1.8	1.4	NA	0.6	0.5	0.9	-0.1	-0.2	0.6	-0.3	0.9	0.0	2.9	1.2	1.3	NA	0.7	1.0	0.2
BST2	AA485371	0.000	0.7	2.0	2.0	0.5	-0.8	0.4	-0.2	1.6	1.9	0.4	-0.2	0.3	1.2	1.5	1.0	1.8	0.5	1.0	0.4	0.6	0.7	0.0
MARCKSL1	AA961735	0.000	0.9	1.5	1.7	0.6	0.5	1.5	1.7	1.0	1.6	1.0	0.5	0.3	0.1	1.4	NA	1.5	-1.3	-0.4	0.1	1.4	0.6	0.3
PKM2	R75662	0.000	1.4	0.3	1.0	0.6	0.6	0.4	0.1	0.2	1.8	0.4	1.0	0.5	0.6	1.2	0.5	1.1	0.9	0.4	0.8	0.7	1.3	1.6
DKFZP564O0463	A1371514	0.000	0.5	0.8	1.0	1.3	0.6	0.6	1.3	1.4	0.6	0.7	1.0	0.5	1.1	0.2	1.0	0.6	0.6	0.4	0.6	0.7	1.3	0.5
PSMD12	AA497132	0.000	0.9	2.2	0.8	1.2	0.7	0.1	0.9	1.9	1.6	0.4	0.2	-0.7	1.2	0.8	0.6	1.1	1.8	-1.5	0.4	0.6	1.0	1.1
DST	H44784	0.000	0.7	2.7	2.4	1.7	0.5	0.5	-0.2	1.6	0.3	0.2	0.6	0.7	-0.3	1.0	NA	NA	NA	-0.8	NA	0.2	1.6	0.7
MMP9	T72581	0.000	1.5	0.6	0.8	2.5	0.5	0.8	-0.3	0.4	0.3	1.5	0.5	0.4	-0.2	1.6	0.5	0.5	1.2	0.1	0.7	0.1	2.4	0.9
PLOD1	AA476240	0.000	0.9	0.4	0.6	1.0	0.6	0.6	0.5	0.7	1.2	1.1	0.7	0.7	0.8	0.2	NA	0.9	0.3	1.0	0.6	0.8	1.3	1.6
ITGA2	AA463610	0.000	0.9	0.7	0.4	-1.8	0.5	0.1	0.9	1.4	0.2	0.2	0.9	0.9	1.6	0.9	0.6	1.4	2.1	0.8	1.2	0.6	1.5	1.0
EYA3	AA262504	0.000	0.2	0.3	0.4	0.6	0.3	0.1	1.4	1.6	1.1	0.9	0.8	0.6	1.0	0.5	0.5	1.2	3.1	-0.5	0.5	0.9	1.0	0.6
ENO1	A1001174	0.000	0.8	1.3	1.3	0.6	-0.2	0.9	0.4	0.9	1.4	0.9	1.2	0.4	1.4	0.8	0.8	0.8	0.4	0.6	0.2	0.3	0.7	1.0
MAP4K4	AA417711	0.000	1.0	0.1	0.1	1.2	0.7	0.6	0.4	0.5	1.0	0.3	0.6	0.7	-0.1	0.1	1.1	NA	NA	0.2	4.1	0.5	1.4	0.8
CTSK	R01515	0.000	1.2	0.7	0.8	1.3	0.5	1.2	-0.5	1.3	0.8	0.2	0.8	0.5	0.0	0.2	1.2	1.0	1.8	0.4	0.1	1.2	1.7	0.6
PHF19	A1337373	0.000	0.2	0.3	2.9	2.2	0.7	0.9	0.4	0.1	0.4	0.1	0.2	1.1	0.2	0.0	0.2	2.1	1.4	0.4	0.4	0.9	1.0	0.8
ZNF71	AA147769	0.144	-0.5	NA	1.3	NA	4.2	-0.8	-0.1	0.4	0.5	1.1	0.2	3.0	0.1	-0.1	NA	NA	-1.0	NA	0.0	1.4	2.2	1.0
NA	R60924	0.000	3.4	NA	NA	-0.1	0.7	0.3	0.7	1.8	1.1	0.3	0.4	0.7	0.4	0.6	1.2	-0.2	NA	0.7	1.5	0.2	0.3	
CLNS1A	T52435	0.000	0.4	0.5	0.8	1.2	1.2	0.3	0.2	1.2	0.4	1.1	0.7	1.3	0.6	0.1	0.8	1.9	0.7	0.7	0.6	1.1	0.3	0.6
TFAP2C	AA399334	0.000	-0.2	0.9	3.2	2.7	0.4	-0.2	1.7	2.1	0.5	0.1	0.4	0.5	-0.3	0.8	NA	0.2	NA	0.9	-0.3	0.4	0.8	0.7
LCN2	N79823	0.000	2.1	1.3	0.9	2.3	0.1	1.0	-0.4	0.0	-0.1	2.1	0.5	-0.3	0.0	2.4	0.9	0.5	0.5	0.2	0.2	0.4	1.7	0.3
HSBP1	AA490864	0.000	0.4	0.7	0.3	0.9	0.9	0.9	1.1	0.9	1.3	1.2	0.7	0.6	1.7	0.1	1.2	0.0	0.8	0.4	0.5	0.7	0.8	0.9
SLC1A3	AA453823	0.000	-0.5	0.3	NA	0.4	0.2	-0.5	0.3	2.1	1.3	-0.3	0.9	0.3	0.2	2.0	1.1	3.7	NA	-0.6	NA	1.1	1.4	1.1
TWIST1	A1220198	0.000	0.8	-0.4	NA	1.4	0.6	NA	NA	NA	1.4	0.2	1.9	0.4	1.1	NA	-0.8	NA	NA	NA	NA	1.5	1.6	0.1
PHYHD1	AA995181	0.083	1.0	NA	-0.1	0.5	0.4	1.9	0.0	-0.1	0.4	1.0	0.1	0.1	0.1	0.2	0.9	0.7	5.2	2.2	0.7	0.2	0.4	0.2
NA	R67336	0.999	-0.7	NA	0.1	0.1	0.6	1.0	1.8	-0.6	1.7	NA	-0.3	0.8	0.0	0.6	-0.3	0.1	7.3	-0.1	1.3	0.4	0.7	0.7
DKFZp564i1922	AA464691	0.000	1.2	0.6	0.0	0.5	0.5	1.3	-0.1	0.9	0.3	-0.2	1.7	0.9	-0.2	0.2	1.5	0.9	1.8	-0.3	1.7	1.2	1.7	0.3
IL1B	AA150507	0.000	0.3	0.4	NA	2.1	0.6	-0.5	0.0	0.2	0.3	0.8	1.6	1.0	0.4	0.6	NA	NA	NA	1.4	0.3	0.7	1.3	2.0
TTYH3	N78076	0.000	0.3	-0.1	0.7	2.8	1.2	0.4	0.5	1.0	0.7	-0.3	0.3	1.1	-0.6	0.6	0.6	2.1	1.3	1.1	1.0	0.9	0.5	0.5
PP	H50229	0.000	0.6	1.0	0.4	0.3	0.4	0.5	1.1	1.9	0.4	2.2	0.6	0.6	1.6	0.9	0.8	0.1	0.7	0.6	1.1	-0.3	0.7	0.5
FABP1	T53220	0.000	0.8	1.2	1.8	1.0	-0.4	0.2	-0.1	0.9	1.4	-0.3	-0.2	0.3	0.5	1.9	1.3	-0.1	1.2	0.5	1.3	0.5	1.5	1.2
RAB31	AA432084	0.000	0.7	0.9	1.1	0.9	0.0	0.5	0.0	0.8	0.6	0.2	0.8	0.1	0.2	1.6	1.6	0.1	1.6	-0.2	0.9	1.1	2.3	0.6
TUBB	H37989	0.000	-0.1	0.7	0.4	0.7	0.7	1.1	0.8	1.3	1.1	1.3	1.3	0.1	1.1	0.8	0.6	0.9	0.8	NA	0.7	0.5	0.3	0.5
ANXA5	AA451895	0.000	1.3	0.7	0.2	0.2	0.3	1.0	0.3	1.1	1.0	1.2	0.6	-0.1	0.7	2.0	0.7	0.9	0.8	0.2	0.7	0.3	1.1	1.2
AIM2	AA458912	0.057	-0.5	0.2	NA	3.7	1.1	0.1	-0.2	NA	0.8	0.0	-0.1	1.8	0.3	0.2	NA	-0.3	NA	0.2	1.1	0.7	3.5	0.8
OCRL	T63362	0.000	2.0	0.3	0.4	0.2	0.3	0.6	-0.1	1.7	0.2	-0.1	2.1	0.2	-0.5	0.3	0.9	1.5	1.9	0.8	0.9	0.9	1.6	0.0
THOC3	AA599058	0.000	0.6	0.5	-0.5	-0.6	1.0	0.9	0.8	0.5	0.6	1.4	1.1	0.6	2.0	0.1	0.7	1.6	3.4	-0.2	0.7	0.8	-0.4	0.8
ALCAM	R13558	0.000	0.3	1.0	1.0	0.6	0.3	NA	-0.2	0.3	1.4	1.1	1.0	-0.7	3.5	0.5	2.3	0.3	1.3	NA	0.4	0.1	-0.2	0.5
IL1RL1	AA128153	0.000	0.9	0.6	0.3	0.3	0.4	0.5	0.8	1.1	0.8	0.9	0.6	0.2	1.4	0.8	2.2	1.1	1.5	0.1	-0.1	0.4	0.7	0.8
HOP	AA504137	0.000	0.3	-0.1	0.1	2.1	0.1	-0.1	-0.3	0.6	0.7	-0.1	0.7	-0.1	0.4	0.6	1.3	0.7	2.6	0.5	1.8	1.1	2.8	0.5

LYN	R83837	0.000	1.0	1.1	1.0	0.3	-0.1	1.2	0.1	0.5	0.4	0.4	0.3	0.0	0.5	1.4	2.3	0.0	1.8	0.1	1.3	0.6	1.4	0.6
RASAL2	AA043653	0.106	0.2	0.0	-0.1	1.6	0.6	-0.4	0.6	0.1	0.2	-0.1	0.1	0.2	NA	-0.1	3.3	3.5	1.1	NA	0.8	-0.2	2.8	0.7
RANBP5	AA598595	0.000	0.8	0.8	0.9	0.7	0.9	0.4	1.7	2.3	0.2	1.1	0.9	0.1	0.7	0.1	0.5	0.5	1.1	0.5	1.0	0.2	0.6	0.1
PLAGL2	AA704187	0.000	0.9	-0.4	-0.1	1.2	1.5	2.6	1.0	1.5	-0.4	-0.2	-0.1	1.9	-0.6	0.2	NA	2.8	NA	1.1	-0.3	1.2	0.2	0.7
PPAP2C	R34980	0.000	0.8	0.0	-0.2	1.4	1.1	1.7	1.1	-0.1	0.4	1.0	0.0	1.5	0.3	0.4	1.6	0.7	1.6	0.4	0.7	0.3	0.6	0.9
APRT	AA598510	0.000	0.7	1.1	0.9	1.8	1.1	0.7	0.3	-0.1	1.4	0.9	0.0	0.7	1.3	0.3	1.3	-0.1	0.4	1.0	0.7	0.6	0.6	0.4
DNAJB1	AA435948	0.000	0.3	0.4	0.8	1.4	0.8	1.1	1.6	0.5	1.0	0.9	-0.2	0.8	0.6	0.7	0.4	1.0	1.9	0.4	0.9	0.2	-0.1	0.6
ZDHHC9	AA293441	0.000	0.6	0.3	0.2	0.8	2.2	2.1	0.6	0.6	-0.2	0.1	0.2	1.3	0.2	0.1	2.3	0.4	2.9	0.3	-0.1	0.5	0.2	0.2
SERPINB2	T49159	0.000	0.5	0.6	0.7	1.1	-0.6	0.8	1.2	0.5	1.2	1.1	0.1	-0.3	1.0	1.2	1.3	0.7	1.0	0.4	0.7	0.6	1.1	1.0
ASPN	T60482	0.000	0.7	0.7	NA	0.5	0.6	0.1	0.4	1.1	0.9	0.2	1.2	0.0	0.9	0.4	3.0	-0.6	2.3	0.3	0.4	1.4	0.9	0.0
CCT3	H27534	0.000	0.2	0.0	0.2	1.6	1.2	0.9	0.9	0.6	1.2	0.4	0.4	0.7	0.5	0.6	0.4	1.1	0.6	0.7	0.9	1.0	1.0	0.9
SMARCC1	AA133216	0.000	0.0	0.8	0.4	0.9	0.8	1.1	1.6	1.0	0.4	0.9	1.1	0.8	1.2	0.6	0.3	-0.4	NA	1.7	0.3	0.9	0.3	0.4
GLO1	AA136710	0.000	0.7	1.1	1.2	0.3	0.7	0.6	1.8	2.2	1.0	1.2	1.0	0.3	1.7	0.3	0.2	0.2	0.2	-0.2	0.1	0.8	0.3	0.3
VKORC1	AA625788	0.000	0.2	0.3	NA	1.0	1.1	0.4	0.7	0.6	1.2	0.4	0.6	0.7	0.4	0.8	0.8	0.5	0.9	0.4	0.6	1.1	1.3	1.3
CDCA7	N45440	0.000	0.7	0.6	1.1	NA	2.1	-0.3	0.6	1.7	0.1	0.8	0.3	1.3	0.5	0.0	2.0	0.8	-0.2	0.3	1.0	0.6	0.9	0.3
CSMD1	N33274	0.000	0.9	1.2	1.5	0.5	0.9	0.6	1.0	1.8	0.9	1.2	1.1	0.1	1.2	0.4	0.3	0.7	0.3	-0.4	0.2	0.6	0.3	0.6
CRTAM	BX118379	0.282	-0.9	NA	NA	NA	2.9	NA	0.7	-2.0	0.3	-0.3	0.5	NA	0.0	1.1	NA	NA	NA	0.5	1.3	-0.2	NA	5.3
CTSB	AA598950	0.000	0.7	0.8	0.7	0.2	-0.3	0.6	-0.3	0.5	1.1	0.9	1.1	0.3	0.7	1.3	0.8	0.6	1.2	0.2	1.9	0.5	1.1	1.3
RRAD	W84445	0.000	1.2	NA	NA	0.9	0.7	2.5	-0.4	0.4	0.5	0.0	0.4	0.0	0.4	0.7	1.7	0.5	NA	0.1	2.1	0.4	0.8	0.4
OAS2	R34568	0.390	-0.1	1.8	1.8	1.6	-2.0	-1.2	-1.8	1.5	1.7	0.2	-0.9	-0.5	-1.0	1.8	NA	NA	0.1	3.5	2.0	3.4	1.7	
TNFSF11	AA504211	0.000	1.4	0.3	1.2	1.3	1.0	NA	-0.2	1.0	0.5	0.2	0.8	0.5	-0.8	0.1	NA	0.1	1.6	1.1	1.5	1.2	0.6	
VBP1	AA478108	0.000	0.3	0.4	1.3	0.6	0.9	0.4	1.1	2.2	1.0	0.6	0.3	0.7	1.6	0.2	1.4	0.7	-0.4	0.1	0.3	0.0	1.4	0.6
ENO1	AA708342	0.000	0.9	1.4	0.5	0.4	-0.3	0.7	0.7	0.5	1.8	0.5	1.2	0.1	1.8	0.9	0.2	0.3	-0.2	0.9	0.3	0.6	1.1	1.5
LIG4	R54358	0.000	0.3	0.7	1.0	1.2	0.6	0.4	1.1	1.0	0.5	1.0	0.2	0.1	1.7	0.2	1.1	0.4	0.9	-0.1	0.5	0.5	1.8	0.6
GGH	AA455800	0.031	-0.7	1.2	1.5	2.0	0.2	2.1	1.7	0.6	1.0	1.8	0.6	0.4	1.9	-1.4	3.0	-0.4	0.7	-0.3	0.4	0.2	-0.6	0.0
GARS	AA629909	0.000	-0.5	0.2	0.3	1.2	0.4	1.2	0.4	0.8	0.7	0.2	0.8	-0.1	0.6	0.2	1.3	0.8	1.9	0.7	0.9	1.3	1.1	1.4
RFC3	H94617	0.000	-0.2	1.0	0.9	1.9	1.5	0.0	1.2	1.7	0.5	0.7	0.3	0.9	0.4	0.2	1.4	0.0	0.5	0.3	0.2	0.3	1.1	0.6
POLR2F	AA418689	0.000	0.5	0.1	0.8	1.4	1.1	0.4	0.8	0.6	0.9	1.1	0.6	0.6	1.3	0.7	0.4	-0.2	NA	0.5	0.6	1.0	1.2	0.8
IGHG3	AA663981	0.508	0.5	2.0	1.7	0.9	1.1	2.5	-1.3	0.5	0.4	0.3	-0.7	0.5	-1.5	4.5	0.0	-2.1	0.9	0.2	2.2	2.0	-0.6	1.5
ZNF313	AA045074	0.000	0.9	-0.5	0.9	1.1	1.0	1.1	1.6	0.7	0.6	0.7	-0.1	0.7	2.5	0.2	0.9	1.1	-0.1	0.8	0.3	0.4	-0.2	0.5
PSAT1	A1015679	0.000	0.5	0.5	0.2	1.1	1.3	0.2	0.7	0.3	1.0	0.3	0.3	1.2	0.7	0.4	0.8	0.0	0.7	0.2	0.1	1.2	1.1	2.4
HSPD1	R26234	0.000	0.9	0.6	NA	2.2	0.0	0.8	-0.1	-0.3	-0.1	0.3	-0.1	0.5	0.8	-0.2	1.3	0.2	NA	2.5	2.1	0.5	0.9	0.8
LY6E	AA865464	0.000	1.3	0.9	0.4	1.3	-0.2	0.9	0.6	1.2	0.9	0.1	1.5	-0.3	0.3	1.8	1.4	0.0	1.0	0.3	0.8	1.3	-0.6	
KCNQ2	H51461	0.373	0.4	0.0	NA	NA	1.0	-0.3	0.2	0.0	0.4	0.6	0.0	0.8	-0.2	-0.1	NA	NA	6.5	0.4	0.8	0.3	0.8	0.9
SLC2A3	AA406551	0.000	0.1	-0.1	0.3	0.1	0.0	1.0	0.4	0.9	1.6	1.2	1.6	0.7	-0.1	0.5	0.6	1.0	2.3	NA	0.6	0.8	0.9	0.5
ZNF143	H98086	0.797	-0.4	2.0	NA	NA	1.6	0.0	0.0	0.4	1.3	-2.9	0.3	5.7	0.3	0.1	NA	NA	NA	NA	NA	1.1	NA	0.4
D2S448	A1356709	0.208	0.6	-0.7	NA	0.6	-0.1	0.2	0.0	0.7	0.5	-0.6	1.8	0.3	-0.7	-1.3	1.6	5.0	1.4	0.3	1.1	1.0	1.9	0.8
PSMD14	N67573	0.000	0.3	1.3	0.6	0.8	0.2	0.7	0.5	-0.5	1.6	1.0	1.1	0.0	1.8	0.2	1.2	0.8	1.0	0.1	0.5	0.4	0.7	1.0
CCS	N30404	0.000	-0.3	0.3	0.4	0.8	1.2	1.0	0.7	0.8	0.6	0.4	0.3	0.8	0.5	0.3	0.6	1.2	1.2	NA	0.1	0.3	2.0	1.2
FLJ11712	N22307	0.166	0.4	0.1	0.4	1.8	1.1	-1.6	0.4	0.8	0.4	0.0	NA	0.5	0.1	-0.1	3.6	3.4	-1.0	1.9	-0.1	0.8	0.9	0.7
LIG3	AA029241	0.000	0.0	0.0	1.3	0.8	1.2	-0.3	0.3	1.0	1.0	0.4	NA	0.4	-0.2	0.5	1.5	0.4	3.2	NA	0.0	0.7	0.7	1.0

IL13RA2	R52796	0.000	0.1	NA	0.8	NA	1.4	0.1	0.5	1.0	0.4	0.9	-0.4	0.1	3.1	0.4	0.2	1.6	NA	0.1	-0.5	1.2	NA	1.4
TNFRSF11B	AA599841	0.321	0.2	0.0	0.0	2.2	1.1	-0.5	0.3	0.3	0.2	1.3	0.6	0.1	0.1	-0.9	1.2	0.0	5.9	0.0	0.8	0.7	1.0	0.4
MAGEF1	AA425302	0.000	0.1	1.5	0.7	0.8	0.7	NA	-0.2	0.4	0.7	1.1	0.5	-0.4	1.2	0.6	NA	0.9	2.8	0.2	0.6	0.9	-0.1	1.0
NA	N73448	0.000	0.0	1.3	0.7	0.6	1.8	-0.2	1.1	0.4	0.3	1.6	0.4	0.7	1.4	0.7	0.7	-0.1	0.3	1.1	0.7	0.8	0.6	0.1
TNFRSF18	AI214481	0.000	2.4	0.7	1.3	-0.6	0.7	0.2	0.3	1.5	0.6	2.0	0.1	0.2	1.2	0.2	NA	1.2	0.7	0.3	0.3	0.3	0.7	0.1
IGF2	N74623	0.081	-0.3	0.6	NA	0.4	4.6	NA	0.0	-0.1	0.3	0.8	0.9	-0.3	-0.4	0.9	2.0	NA	NA	0.6	1.0	0.5	0.6	0.1
RUNX1	AA425238	0.000	0.7	0.0	-0.7	1.6	0.5	1.0	0.0	0.6	0.7	0.0	0.8	0.4	0.3	0.6	2.0	1.1	1.3	0.3	1.3	0.7	1.7	0.1
NA	AA455521	0.000	0.1	1.0	1.0	2.2	0.5	0.8	0.3	0.6	0.6	1.3	0.0	0.5	0.6	0.4	1.0	1.7	-0.9	0.8	NA	1.0	0.8	0.1
CDC45	W00895	0.000	0.8	-0.3	1.7	1.8	0.3	0.8	0.3	0.4	0.7	0.1	0.1	0.8	0.4	0.2	0.9	1.8	1.0	0.1	0.4	1.0	0.8	0.1
PCDHA6	AA172188	0.797	0.6	0.2	NA	2.2	NA	0.6	-0.4	NA	1.2	-0.1	-0.5	0.6	NA	-0.1	NA	-1.4	4.3	1.8	0.5	1.4	0.5	0.1
BRIX	AA454607	0.000	0.6	0.4	1.0	1.3	0.5	0.6	1.2	0.7	1.1	1.2	0.7	0.4	1.0	0.5	0.0	0.5	0.2	0.0	0.4	1.0	0.8	0.7
IARS	AA410636	0.000	1.0	0.9	0.9	0.4	0.5	0.8	0.4	0.2	0.4	0.3	1.0	0.3	1.4	0.6	0.9	0.2	1.7	-0.1	0.4	0.8	0.7	1.1
ZNF76	AA626012	0.000	-0.1	0.5	0.1	0.5	0.7	-0.3	0.9	1.6	0.4	0.6	2.0	0.4	0.0	0.8	1.3	0.4	0.9	0.2	0.6	1.0	2.0	0.5
SLCO2A1	AA037014	0.000	1.4	0.3	2.4	2.0	0.2	0.0	-0.1	0.7	0.5	0.4	1.6	0.3	-0.4	-0.2	0.1	0.2	NA	0.2	0.3	1.7	1.8	0.8
PGF	AA130714	0.000	0.0	0.4	NA	1.5	0.5	-0.1	0.1	0.4	0.1	0.8	0.0	0.8	0.7	1.4	3.2	NA	0.9	0.7	0.4	0.5	0.8	0.3
DCTD	H68309	0.000	0.7	-0.2	0.4	0.7	1.1	1.6	0.5	0.4	0.7	0.9	0.3	0.9	1.7	1.2	-0.1	0.6	0.1	0.7	0.7	0.7	0.5	0.6
CD36	W93331	0.000	1.4	0.8	0.8	0.9	0.2	1.0	0.9	2.1	0.6	0.3	0.1	0.8	0.7	-0.5	1.3	1.7	0.7	0.3	0.5	-0.6	0.2	0.4
UCHL3	N27190	0.000	0.4	0.5	0.6	0.8	1.0	0.3	1.2	1.6	0.7	0.9	0.6	0.4	1.2	-0.2	1.1	-0.1	1.5	0.3	1.3	-0.2	0.9	0.2
AP1M1	AA644232	0.032	2.4	0.1	1.0	0.8	0.5	0.3	-0.3	0.2	0.1	NA	0.5	0.6	-0.4	0.0	0.1	1.9	3.7	1.1	0.2	0.3	0.2	0.5
CXCR4	T62491	0.000	2.2	0.0	0.3	0.9	0.9	0.4	-0.5	0.6	0.4	-0.5	1.5	0.4	-0.4	0.4	1.0	0.6	1.3	0.5	0.8	1.2	2.6	0.2
CML66	N32587	0.000	0.4	1.0	1.0	0.7	0.0	1.6	0.7	0.7	0.3	0.7	0.1	0.1	1.2	0.1	1.7	2.6	-0.7	0.7	0.6	0.5	0.6	0.2
BIRC5	AA460685	0.000	0.2	0.7	0.8	2.5	0.3	0.3	0.3	0.0	1.3	1.0	0.2	0.5	0.6	0.1	0.4	0.9	0.3	0.9	0.9	1.2	0.4	0.4
NUP155	R21785	0.000	-0.3	0.7	1.2	0.4	-0.3	-0.2	0.3	2.0	0.6	0.7	0.0	-0.1	2.0	0.4	0.7	0.1	3.1	1.0	NA	1.1	0.2	0.5
TTN	H97920	0.031	-0.2	0.3	-0.4	NA	2.9	1.0	0.7	-0.3	0.2	-0.1	0.3	0.4	0.1	-0.1	NA	NA	NA	NA	3.9	0.6	NA	1.3
PRDX1	AA775803	0.000	0.4	0.2	0.0	0.3	0.2	1.1	0.9	0.4	1.2	0.4	0.6	0.9	0.6	0.8	0.9	1.1	0.3	0.8	0.6	0.9	1.2	1.0
DC13	AA181314	0.034	1.8	0.9	0.5	NA	0.2	1.5	0.3	0.1	1.0	0.9	0.1	0.2	2.7	0.5	2.4	-0.6	-0.5	0.4	1.7	-1.1	0.1	0.9
UBE4B	A082562	0.000	0.3	1.5	1.1	0.9	-0.7	0.5	0.2	1.1	1.7	0.9	-0.1	0.3	-0.1	2.1	0.6	0.7	0.4	0.6	0.3	0.6	1.1	0.1
CDK5RAP2	A040627	0.000	0.5	0.0	0.5	3.2	0.8	0.5	-0.1	0.6	0.7	0.1	-0.1	1.1	0.4	-0.3	-0.3	1.2	NA	0.9	0.4	0.5	2.6	0.7
TUBA3	AA865469	0.000	0.2	0.9	-0.8	0.9	-0.1	0.9	1.1	1.1	1.5	0.5	0.8	-0.1	1.3	0.8	NA	NA	0.7	0.4	0.9	0.5	0.8	1.0
ABHD2	AA454207	0.000	1.2	NA	0.8	2.0	0.3	-0.5	0.0	0.8	0.5	1.6	-0.3	0.0	0.0	1.1	NA	1.7	1.0	0.3	-0.2	0.1	1.8	0.8
PSMB9	AA862434	0.000	0.0	1.5	1.2	0.4	-0.2	1.3	0.8	1.1	1.9	0.7	-0.3	0.7	0.7	1.0	NA	0.1	0.1	0.4	0.3	0.8	0.8	0.3
UBE2L6	AA292074	0.000	0.3	1.3	1.0	0.9	-0.8	1.0	0.4	1.3	1.6	0.8	-0.2	0.3	0.1	1.6	0.2	1.2	0.5	-0.1	0.5	0.3	0.8	1.3
RABGGTB	AA456028	0.000	0.4	1.1	1.1	0.4	0.6	0.4	0.3	0.3	0.3	0.2	0.8	0.4	1.0	0.1	0.2	2.9	1.0	-0.2	NA	1.1	0.5	0.7
NP	AA430382	0.000	0.2	0.7	0.8	0.9	0.3	0.3	1.2	0.5	0.6	0.6	0.7	-0.3	1.8	0.5	1.2	0.1	0.4	0.2	0.9	0.9	0.9	1.2
ITGB3	AA037229	0.000	1.2	0.1	0.2	1.1	0.1	-0.2	-0.1	0.7	1.0	0.4	1.1	1.2	-0.1	-0.3	1.3	1.6	1.3	0.1	-0.3	0.5	1.4	2.1
CDKN3	AA284072	0.000	0.5	NA	NA	1.4	0.7	0.4	0.3	0.8	1.0	-0.7	0.1	0.4	NA	0.5	1.3	NA	NA	NA	1.4	0.5	1.1	0.7
BUB3	H38804	0.000	0.7	0.9	2.8	0.2	0.2	-0.2	0.6	0.8	0.4	0.7	0.4	0.1	1.0	0.5	1.2	0.9	0.9	0.6	0.1	0.7	0.1	0.7
PDCD5	AA156940	0.000	0.0	0.7	0.8	1.1	1.0	0.8	1.0	0.3	0.8	1.2	0.5	0.7	NA	0.0	0.5	0.9	0.9	-0.2	0.6	1.2	0.5	0.4
CDC25B	AA448755	0.000	0.9	-0.1	0.2	2.2	0.3	0.7	0.5	0.1	1.7	0.5	-0.1	0.9	-0.1	0.8	0.4	1.7	1.2	0.4	0.2	0.8	1.7	0.4
HDAC2	AA127093	0.147	0.1	0.6	NA	0.6	0.6	-0.1	0.7	-0.1	1.0	0.6	0.4	0.1	0.6	-0.2	0.3	1.4	4.9	-0.2	0.1	NA	0.8	0.7
HGF	R52797	0.000	0.1	0.3	0.0	1.6	0.7	0.0	0.6	-0.1	0.6	0.2	0.7	0.7	1.0	0.6	2.6	0.9	-1.1	1.6	0.4	1.1	0.7	1.0

C9orf16	AA398406	0.000	0.6	1.1	1.2	0.5	0.4	0.3	0.2	0.2	0.8	1.2	0.3	0.4	1.0	0.6	0.8	0.4	2.1	0.6	-0.2	0.6	0.6	0.6
MGC508	AA504202	0.000	0.5	0.1	0.5	1.2	0.6	0.7	0.4	0.3	0.3	0.8	0.6	0.6	0.3	1.0	1.5	1.0	1.0	0.3	1.0	0.3	0.6	0.6
PSMB7	T97171	0.000	0.9	0.6	1.1	0.6	0.7	0.4	0.5	0.0	0.8	0.8	0.2	0.4	1.4	0.4	2.1	0.4	0.5	0.3	0.8	0.5	-0.2	0.9
CDC45L	AA700904	0.000	0.4	0.1	1.0	2.7	0.5	0.0	0.4	-0.1	1.1	0.2	-0.5	0.7	0.5	0.7	0.5	0.6	0.9	0.6	0.4	1.0	NA	2.0
SOC51	AA280137	0.758	0.3	-0.1	-1.0	NA	1.1	-0.6	0.0	0.2	0.0	0.2	-0.3	1.4	0.1	0.7	4.6	-0.5	4.8	0.3	0.2	0.8	0.9	0.5
G3BP	AA449834	0.034	0.1	0.2	1.1	3.3	0.5	-0.1	0.5	2.4	0.5	1.0	0.6	0.4	0.2	-0.4	0.5	0.6	-0.6	0.3	0.2	1.9	0.9	0.2
C9orf26	W79445	0.321	0.4	0.8	1.3	0.3	-0.3	0.1	0.8	0.1	-1.1	1.7	3.3	-0.7	1.7	-1.0	0.4	-1.0	1.5	2.2	1.1	2.8	-0.8	0.5
TUBE1	AA215502	0.000	-0.4	0.0	NA	1.9	1.3	0.3	0.4	0.4	1.0	0.4	0.1	1.3	0.7	0.4	1.1	NA	0.5	1.0	NA	0.4	0.7	NA
SDC4	AA148736	0.000	0.0	0.5	-0.4	0.7	0.5	1.0	1.4	1.1	NA	0.8	0.4	0.4	0.7	0.7	1.0	1.0	0.7	0.8	0.9	0.4	0.4	0.6
VCAM1	H07071	0.000	0.2	0.8	0.9	1.4	0.5	1.1	1.1	1.3	-0.2	1.0	0.6	0.6	0.2	1.1	0.0	0.9	0.7	0.5	0.1	1.2	0.2	-0.1
LOC150223	AA454833	0.000	0.2	0.2	0.7	1.1	1.5	0.4	0.3	0.7	1.0	0.5	0.3	0.4	0.4	0.7	0.2	-0.2	0.7	1.3	0.5	1.0	1.0	1.1
SERPINE1	N75719	0.000	0.7	0.1	0.1	NA	-0.1	0.3	0.3	0.3	0.6	-0.1	1.5	1.0	-0.3	-0.2	1.1	1.1	1.2	0.1	0.7	1.5	3.3	0.2
LATS2	AA431967	0.000	0.5	0.1	0.2	1.2	0.2	0.3	-0.1	1.8	0.3	-0.2	0.2	0.4	-0.6	1.1	1.1	0.4	2.2	0.6	1.3	0.7	1.6	0.6
CRH	R45054	0.186	-0.1	1.5	1.3	1.0	-0.3	1.9	1.6	0.6	0.1	1.8	0.7	0.2	2.8	-1.3	3.1	-0.3	0.7	0.0	0.1	-0.3	-1.0	-0.2
NA	A1018643	0.000	0.3	0.4	1.0	0.8	0.5	0.9	0.5	0.5	0.2	0.4	0.4	0.8	0.6	0.1	1.1	1.5	2.5	-0.1	1.1	0.3	0.1	0.1
RAD23B	AA489678	0.000	0.8	1.5	0.3	0.7	0.6	0.2	0.8	-0.1	1.0	0.4	0.9	0.5	1.3	0.3	0.5	0.1	NA	0.2	0.7	0.9	0.9	0.9
EIF3S2	AA936783	0.000	1.0	0.2	1.1	0.6	0.9	0.4	0.3	0.9	0.8	0.6	0.4	2.7	1.4	-0.1	0.1	0.3	0.8	0.4	0.4	0.0	0.1	0.6
INSIG1	H59620	0.034	1.1	1.3	2.5	0.5	-0.6	0.8	0.4	1.0	0.2	0.0	-0.1	-0.9	0.3	-0.4	1.9	1.0	1.2	0.6	0.4	-0.6	1.7	1.6
IFI30	AA630800	0.032	0.6	0.8	1.6	0.9	-0.5	0.0	-2.0	0.7	1.7	0.4	-0.1	0.3	0.2	1.8	1.6	-0.1	1.7	0.1	1.3	0.3	1.4	1.2
ZNF593	AA033532	0.000	0.4	0.4	0.6	1.5	0.8	1.0	0.4	0.5	0.5	0.7	0.5	0.7	0.7	0.1	1.4	0.1	0.3	0.8	0.4	0.4	0.0	0.1
COL4A2	AA430540	0.000	0.5	0.3	0.0	1.3	0.4	0.8	0.4	0.6	0.2	-0.2	0.7	0.5	-0.1	-0.2	0.7	1.5	0.7	0.7	0.8	1.9	1.5	0.7
GTL3	AA425682	0.000	0.5	-0.4	0.2	1.0	0.8	1.0	0.8	0.4	1.0	1.1	-0.3	0.9	0.8	0.3	0.3	1.3	1.2	0.2	0.8	0.8	0.3	0.8
PUS1	AA496794	0.000	0.6	0.1	1.0	0.6	0.7	0.8	-0.2	0.1	0.1	0.5	0.3	1.1	0.0	0.1	3.1	1.1	1.7	0.3	0.9	0.6	-0.1	0.3
ZNF587	AA719022	0.000	-0.2	0.7	0.2	1.8	0.8	0.8	0.4	1.3	0.2	0.4	0.2	0.9	0.3	-0.2	0.7	1.1	0.4	0.1	0.7	0.8	0.5	0.5
CUL5	AA086475	0.000	0.2	1.0	1.0	1.1	-1.0	0.8	0.3	1.1	1.6	0.8	0.1	0.3	-0.3	1.9	0.3	0.9	0.5	-0.1	0.3	0.7	1.1	1.1
MGC8974	H90225	0.000	1.2	-0.9	0.8	0.4	0.3	0.8	0.5	-0.1	1.1	0.7	-0.3	0.0	1.2	1.2	1.3	2.1	NA	0.4	0.5	0.7	0.2	0.9
DLG5	T77840	0.390	-0.2	0.2	0.9	0.5	0.0	-0.2	1.1	0.4	0.3	-0.3	0.3	0.6	5.7	0.3	1.5	0.6	-0.1	0.3	1.2	0.3	0.1	0.3
KCNE3	A1017405	0.000	1.1	0.5	1.4	1.0	1.2	0.4	0.0	0.9	-1.0	1.1	0.7	0.8	1.1	0.2	-0.1	0.6	1.3	-0.6	0.9	1.1	0.7	0.5
PARP14	T64956	0.000	0.8	1.6	0.8	-0.3	-1.2	0.4	0.2	1.6	0.8	0.3	0.0	0.1	0.1	2.2	1.1	0.7	1.0	0.1	1.1	1.1	0.5	0.5
RPS7	A1675707	0.000	0.2	0.8	0.5	1.1	1.7	0.0	1.0	0.5	0.5	0.7	0.5	0.7	0.9	0.4	0.1	-0.2	0.5	0.7	0.9	1.4	0.5	0.2
CCT5	AA629692	0.000	0.5	0.6	0.6	0.4	0.4	1.4	1.2	0.8	0.8	0.5	0.6	0.0	1.3	0.3	0.5	0.7	0.4	0.3	0.1	0.6	0.6	0.8
FZD6	T68333	0.000	0.7	0.7	0.7	0.3	0.5	0.5	0.6	1.2	0.7	1.5	0.8	0.5	0.6	0.2	0.8	0.9	1.3	-0.5	0.3	1.0	0.2	0.2
NEK6	AA463188	0.000	-0.3	0.1	1.9	2.4	1.0	0.7	-0.2	0.0	0.5	0.0	-0.1	0.8	-0.8	0.1	1.1	1.9	NA	0.3	NA	0.5	1.8	0.6
NA	R06567	0.000	0.7	0.3	1.5	1.9	0.7	0.2	0.1	0.3	1.8	0.2	0.1	0.1	1.3	0.6	-0.2	0.4	0.6	0.0	0.5	0.9	1.2	0.6
DCC1	AA934904	0.000	0.6	0.2	0.5	1.8	0.3	0.3	0.8	1.2	1.7	0.8	-0.3	0.4	-0.3	0.9	0.8	0.5	0.8	0.5	NA	0.0	1.1	0.3
CD99L2	R94508	0.000	-0.5	NA	NA	1.7	0.9	1.6	0.1	0.4	0.8	0.9	0.4	0.6	1.1	0.0	NA	1.3	-0.7	1.3	0.1	0.6	0.8	0.6
FLJ10534	AA432108	0.000	0.8	0.5	1.1	NA	NA	1.7	-0.1	NA	1.0	0.2	0.5	0.5	0.4	0.2	0.7	1.4	1.2	0.6	0.6	0.1	0.0	0.6
NA	AA916728	0.000	0.0	0.5	0.8	1.4	0.6	0.6	0.1	0.3	0.3	1.4	0.3	0.5	0.6	0.8	0.8	1.7	1.2	0.3	0.7	0.8	-0.3	0.3
ANGPTL2	AA704833	0.000	0.8	-0.1	-0.2	1.0	0.2	0.5	0.2	0.5	1.5	0.2	0.3	0.6	0.3	0.0	0.9	1.2	2.3	0.1	0.6	0.7	2.0	0.4
PTK7	AA453789	0.167	0.3	0.1	0.5	1.7	1.3	0.3	1.1	0.2	1.1	0.4	0.5	0.9	0.8	-3.5	2.1	1.4	1.6	0.6	0.7	0.7	0.9	0.0
FHL3	AA460438	0.000	0.4	0.0	0.9	1.1	1.0	0.2	0.1	-0.2	0.6	0.3	-0.2	0.1	-0.1	0.4	0.6	0.3	3.6	0.7	1.5	1.0	0.8	0.6

LEPRE1	AA976184	0.000	2.4	NA	1.0	0.3	0.1	-0.2	0.1	0.1	0.3	1.1	0.7	0.1	1.7	0.1	1.5	NA	1.3	0.6	0.0	0.2	0.8	0.2
CCL4L	H62985	0.000	0.4	0.0	0.8	1.4	-0.5	-0.2	0.2	0.8	0.9	0.2	0.1	0.7	-0.2	2.4	0.6	0.1	2.7	0.0	0.3	0.6	1.3	0.9
POLR1C	AA733038	0.000	0.4	1.1	0.0	1.2	0.3	0.6	0.5	0.5	0.3	1.0	0.8	0.5	1.0	0.4	1.0	-0.9	1.6	0.7	0.5	1.2	0.4	0.6
DHCR7	AI652764	0.083	0.4	-2.7	0.3	2.2	1.2	1.6	0.7	0.7	1.5	0.2	0.2	0.9	0.3	0.0	0.8	0.8	0.2	-0.4	0.6	1.0	1.1	2.0
MSH2	AA219060	0.000	0.9	0.3	1.3	0.1	-0.3	NA	1.3	0.9	1.0	0.8	NA	0.1	1.0	0.5	1.3	0.6	0.9	0.0	0.2	0.4	0.5	0.5
WBSCR14	R10526	0.000	1.3	0.1	0.6	2.6	0.3	0.6	0.1	0.3	0.0	0.2	0.3	1.5	0.2	0.1	1.0	0.8	1.1	0.2	0.3	0.1	1.4	0.2
IL15RA	AA053285	0.103	1.6	0.8	0.0	0.7	0.0	0.9	-0.1	0.3	0.4	0.1	0.0	0.5	NA	0.5	-0.2	4.6	NA	0.4	0.3	0.7	0.3	0.5
ZNF259	AI343293	0.000	0.0	0.4	0.2	1.4	1.1	0.7	0.7	0.5	0.5	0.5	0.3	0.5	1.0	0.9	1.1	1.1	NA	0.1	0.3	0.8	0.5	0.3
NUFIP1	AI653424	0.000	0.5	0.1	0.4	0.7	0.9	0.1	0.8	1.2	0.7	0.2	-0.7	1.1	0.5	0.1	1.3	0.9	0.8	-0.1	1.3	1.3	0.5	0.7
FLJ25952	AA683490	0.000	1.2	0.8	1.0	NA	0.1	0.4	0.5	-0.1	0.2	1.2	0.6	0.5	1.0	-0.3	2.0	1.6	0.3	0.3	0.1	-0.5	1.7	0.2
HRASLS3	AA476543	0.000	1.5	0.1	0.1	0.6	-0.2	2.5	0.1	0.8	1.0	0.7	-0.5	0.6	0.7	1.8	0.0	0.6	0.5	0.2	0.4	0.0	1.6	0.2
TFAP2A	R38133	0.000	0.3	0.4	NA	0.7	0.5	-0.6	0.5	0.0	0.8	0.2	1.7	0.9	0.1	0.4	NA	NA	0.6	0.6	0.9	1.9	1.2	
UBE2M	AA449119	0.000	0.3	0.3	0.4	1.9	0.7	0.2	0.5	0.2	1.2	0.5	0.2	0.8	0.8	0.7	0.2	1.1	-0.2	0.5	0.9	0.8	0.6	0.9
NK4	AA458965	0.034	1.1	0.7	0.7	-0.1	-0.3	2.1	0.2	0.7	1.3	-0.5	-0.4	1.3	0.2	2.0	2.7	0.7	-0.1	0.3	0.0	0.4	0.1	0.7
GABPB2	H91651	0.000	0.5	0.7	1.5	0.8	0.1	1.2	0.9	0.6	0.6	0.6	0.4	0.2	0.5	0.1	0.1	0.9	2.7	-1.1	NA	0.5	0.3	0.6
RAD54B	AA973748	0.000	0.9	NA	-0.4	1.0	0.1	0.1	1.4	0.4	0.5	0.3	-0.5	0.3	1.5	0.5	2.8	0.3	NA	NA	-0.1	0.7	1.2	0.9
PPP1R3C	AA133278	0.147	-0.2	0.4	2.7	3.1	0.3	-0.8	0.1	1.2	0.3	0.1	0.0	-0.5	-0.3	0.2	0.8	1.9	1.0	-0.4	0.4	-0.1	1.4	1.5
ARL7	AA485683	0.060	0.2	-1.0	2.6	1.7	0.5	2.2	0.0	0.2	0.6	0.5	0.3	0.4	0.3	0.3	NA	2.2	1.8	-0.2	-0.3	0.1	0.0	0.2

The q-value (%) set up the lowest False Discovery Rate associated with each selected genes. For every patient, data is reporting, in log 2 scale, expression ratios between tumor biopsies and their normal counterpart. NA: Not Available.

Anexo II

Supplemental Material can be found at:
http://www.mcponline.org/cgi/content/full/M600048-MCP200/DC1

Research

Proteomics-based Validation of Genomic Data

APPLICATIONS IN COLORECTAL CANCER DIAGNOSIS*[§]

Juan Madoz-Gúrpide†§¶, Paula López-Serra†§¶, Jorge Luis Martínez-Torrecuadrada†, Lydia Sánchez**, Luis Lombardía††, and J. Ignacio Casal†§§

Multiple factors are involved in the translation of functional genomic results into proteins for proteome research and target validation on tumoral tissues. In this report, genes were selected by using DNA microarrays on a panel of colorectal cancer (CRC) paired samples. A large number of up-regulated genes in colorectal cancer patients were investigated for cellular location, and those corresponding to membrane or extracellular proteins were used for a non-biased expression in *Escherichia coli*. We investigated different sources of cDNA clones for protein expression as well as the influence of the protein size and the different tags with respect to protein expression levels and solubility in *E. coli*. From 29 selected genes, 21 distinct proteins were finally expressed as soluble proteins with, at least, one different fusion protein. In addition, seven of these potential markers (*ANXA3*, *BMP4*, *LCN2*, *SPARC*, *SPP1*, *MMP7*, and *MMP11*) were tested for antibody production and/or validation. Six of the seven proteins (all except *SPP1*) were confirmed to be overexpressed in colorectal tumoral tissues by using immunoblotting and tissue microarray analysis. Although none of them could be associated to early stages of the tumor, two of them (*LCN2* and *MMP11*) were clearly overexpressed in late Dukes' stages (B and C). This proteomic study reveals novel clues for the assembly of a robust and highly efficient high throughput system for the validation of genomic data. Moreover it illustrates the different difficulties and bottlenecks encountered for performing a quick conversion of genomic results into clinically useful proteins. *Molecular & Cellular Proteomics* 5:1471–1483, 2006.

Colorectal cancer (CRC)¹ is the second most prevalent

From the †Protein Technology, **Histology and Immunohistochemistry, and ††Genomics Units, Biotechnology Programme, Centro Nacional de Investigaciones Oncológicas (CNIO), 28029 Madrid, Spain Received, February 7, 2006, and in revised form, May 5, 2006 Published, MCP Papers in Press, May 29, 2006, DOI 10.1074/mcp.M600048-MCP200

¹The abbreviations used are: CRC, colorectal cancer; MGC, Mammalian Gene Collection; MBP, maltose-binding protein; CNIO, Centro Nacional de Investigaciones Oncológicas; SAM, significance analysis of microarrays; HRP, horseradish peroxidase; BMP, bone morphogenetic protein; LCN2, lipocalin 2; MMP, matrix metalloproteinase; ANXA3, annexin A3; SPP1, secreted phosphoprotein 1 (osteopontin); SPARC, secreted protein, acidic and rich in cysteine (osteonectin); THBS2, thrombospondin 2; HIG2, hypoxia-inducible gene 2; AGT, angiotensinogen precursor; PLAU, plasminogen activator.

cancer in the western world. CRC develops over decades and involves multiple genetic events. From a genetic point of view, CRC is one of the best studied solid tumors (1). However, it is paradoxical that despite this level of knowledge and the existence of good screening procedures CRC continues to be a major cause of mortality in the developed countries. So the implementation of simpler, non-invasive methods for the early detection of CRC is necessary. These methods should be based on new biomarkers, preferably proteins or antibodies detectable in serum or plasma.

Traditional methods of identifying novel targets involved in cancer progression were based on studies of individual genes. Now the use of DNA microarrays permits the analysis of the expression of tens of thousands of genes simultaneously and rapidly (2, 3). Microarray analysis has been used for gene expression analysis of different neoplasms (4, 5) including CRC (6–9). However, only a few studies have pursued further insight into the function and/or importance of individual genes and their application to the proteome research of a tumor. Some of these genes have been proposed as candidate cancer biomarkers (10–12). More recently a number of proteomic studies have also addressed the identification of potential targets in CRC (13–15).

As a part of a comprehensive approach to study the proteome of CRC and to identify new biomarkers, we investigated the feasibility of expressing soluble proteins corresponding to up-regulated genes in cancer patients with surgically resected colon polyps and tumors. We used cDNA microarrays (CNIO Oncochip) (16) to identify differentially expressed genes in malignant versus normal samples isolated from individual patients with CRC.

However, to carry out these studies, it is necessary to design and standardize high throughput methods and strategies to translate genes into proteins (or protein fragments) and subsequently generate antibodies in a quick manner. To retrieve the necessary genes we chose two possible routes. The first alternative was the cDNA clones from the IMAGE Consortium that were printed on the Oncochip (16). The second was the full-length cDNA clones either from the Mammalian Gene Collection (MGC) (17, 18) or from other sources (i.e. donations, etc.). Regarding the expression system, *Escherichia coli* is still the most suitable host for massive production of recombinant proteins in a high throughput approach because of its fast growth, high production rates, and cost effectiveness. It is necessary to obtain as many soluble pro-

Validation of Genomic Targets in Colorectal Cancer

teins as possible because insoluble proteins are generally inadequate for antibody production, crystallization, or functionality studies in general. For this reason, we selected three fusion tags: His₆, GST, and MBP; the latter two have been described previously as very efficient solubilizing agents in selected subsets of human proteins (19, 20). A limitation to the use of high throughput expression procedures is the use of conventional cloning procedures based on restriction and ligation of DNA fragments, which are not too efficient and are excessively time-consuming. Here we propose the use of recombinational cloning systems such as the Gateway cloning technology (21, 22). This system works in two recombination steps: the first step is the creation of an entry clone, and the second step is the subsequent generation of multiple destination vectors. These steps are usually very efficient and can be easily made in parallel for many different genes.

Another limitation in proteome research and target validation is the poor availability of antibodies against the gene products that are being identified in massive numbers either by DNA microarray analysis or by proteomic technologies such as two-dimensional PAGE or LC-MS/MS. In many cases there is no possibility to rapidly assess the actual value of new markers because the number of potential protein candidates exceeds by far that of existing antibodies. The use of antibodies for tissue profiling (*i.e.* in tissue microarrays) allows for a fast approach to generate protein expression data for normal and disease tissues on large numbers of individual patients (23). Only antibodies possess the sufficient specificity for a proper detection of the target proteins. As a result, the validation of some of those proteins that might become biomarkers for screening or prognostic value is considerably delayed. Moreover specific antibodies are needed for numerous functional assays, including ELISAs, localization studies, and "pulldown" experiments (24). Mouse-derived monoclonal antibodies continue to be the affinity reagent of choice in proteomic analyses, but their production remains restricted by antigen quality, high tissue culture requirements, and low throughput screening methods. Rabbit polyclonal antibodies are a faster and cheaper alternative, but the quality and specificity depend largely of the quality of the antigen. Other methods such as phage display that may become useful alternatives are currently being tested.

In this report, we describe the translation of potential gene targets into soluble proteins by using a recombinational cloning system and their evaluation by immunoblotting and tissue microarray. Important bottlenecks and limitations will be described in protein production and purification as well as in antibody production. As a proof of the usefulness of this approach for selecting potential biomarkers for CRC, seven proteins (annexin A3 (ANXA3), lipocalin 2 (LCN2), bone morphogenetic protein 4 (BMP4), osteonectin (secreted protein, acidic and rich in cysteine (SPARC)), osteopontin (secreted phosphoprotein 1 (SPP1)), and matrix metalloproteases 7 and

TABLE I
Clinical data of samples included for gene expression analysis

Biopsy no.	Tumor no.	Normal no.	Age	Sex ^a	Tumor location ^b	Dukes' stage
			yr			
1010118	36T	35N	78	M	AC	C
1010283	80T	78N	78	M	AC	C
1010286	92T	71N	45	F	DC	C
1010292	81T	72N	63	M	DC	C
1010320	77T	90N	71	M	AC	C
1010322	96T	95N	72	F	DC	C
1010414	107T	106N	81	F	DC	C
1010433	84T	87N	56	F	AC	C
1010445	83T	85N	79	F	AC + DC	C
1010455	109T	108N	67	M	DC	C
1010457	111T	110N	81	F	AC	C
3011002	1T	15N	68	M	DC	B
3011003	59T	37N		M	AC	B
3011011	20T	26N	80	F	AC	A
3011021	30T	29N	40	F	AC	B
5011390	62T	41N	75	M	AC	A
5011392	48T	47N	71	F	DC	B
6010532	40T	39N	87	F	AC	A
6010533	46T	45N	54	F	AC	C
9010070	54T	53N	77	M	AC	C
9010085	56T	55N	80	M	AC	B
9010086	58T	57N	78	F	DC	C

^a M, male; F, female.

^b AC, ascending colon; DC, descending colon.

11 (MMP7 and MMP11) were characterized for CRC diagnosis and characterization.

EXPERIMENTAL PROCEDURES

DNA Microarray Experiments—Tissue samples from 22 CRC patients were provided by the Tissue Bank Network of the CNIO in collaboration with the following hospitals in Spain: Virgen de la Salud (Toledo), Cartagena (Murcia), Puerta de Hierro (Madrid), Clínico de San Carlos (Madrid), Ramón y Cajal (Madrid), and Alcorcón (Madrid). These frozen samples consisted of pairs of tumoral and normal mucosa sections. Clinical and pathologic data of the samples are shown in Table I. Total RNA was isolated from 50–100 mg of frozen biopsies as described previously (16, 25–27). To generate fluorescent cDNA, 30 µg of tumoral or normal RNA samples were labeled, respectively, with Cy5-dUTP and Cy3-dUTP (GE Healthcare). Hybridized slides were scanned with the Agilent G2565BA Microarray Scanner System, and images were then quantified using the GenePix Pro 4.0 application (Axon Instruments Inc.). After raw data normalization with the "Lowess" routine (28, 29), multiple testing to detect differentially expressed genes was carried out using the significance analysis of microarrays (SAM) method (30). One class response analysis was carried out, allowing for 2,000 random permutations.

Prediction of Surface Extracellular Proteins—Different bioinformatics algorithms were used for protein selection and classification. The Sosui system (classification and secondary structure prediction of membrane proteins, sosui.proteome.bio.tuat.ac.jp) was used to discriminate membrane from soluble proteins. To perform gene ontology analysis we used the Gene Ontology website (www.geneontology.org/) that uses controlled vocabularies (ontologies) describing gene products in terms of their associated biological processes, cellular components, and molecular functions in a species-independent manner. Our search was achieved by tracking concepts such as "extra-

Validation of Genomic Targets in Colorectal Cancer

cellular, growth factor, membrane, and receptor."

cDNA Clones—Twenty-nine up-regulated genes were selected to express their corresponding protein products. Three sources of cDNA clones were used as starting material for cloning and expression of the activated genes: initially cDNA clones for the whole set of targets were obtained from the 40 K collection from the IMAGE Consortium (Research Genetics). The requisite for attempting expression was to contain at least 30% of the ORF. Another 24 full-length cDNA clones were obtained from the MGC and purchased from MRC Geneservices. SULF1 and thrombospondin 2 (THBS2) cDNA clones were a kind gift from Dr. Rosen (University of California San Francisco) and Dr. Bernstein (University of Washington), respectively. For three clones (COL5A2, ITGA2, and SLC2A1) we were unable to obtain a full-length sequence cDNA clone. All the clones were resequenced to verify the inserted sequence by using an automatic DNA sequencer.

Primer Design and Cloning in the Gateway System—For amplification of the coding regions, oligonucleotides were designed in such a way that the regions coding for the leader sequences of the proteins were removed to facilitate expression in *E. coli*. For identification of signal peptides we used SMART (Simple Modular Architecture Research Tool), which allows for a rapid identification and annotation of signaling domain sequences (smart.embl-heidelberg.de). For a fast cloning of the genes into the Gateway system we used the directional pENTR/D-TOPO vector (Invitrogen). 5'-Oligonucleotides were designed to place the TOPO leader sequence (CACC) immediately upstream of the ATG initial codon, resulting in primers with an average size of 28 bp. 3'-Oligonucleotides were designed either to contain the ORF stop codon in full-length clones or to introduce a novel stop codon for partial fragments. Oligonucleotides were purchased from Sigma Genosys. Coding regions were amplified from the cDNA clones by PCR with Vent DNA polymerase (New England Biolabs). For the cloning, 1 μ l of the PCR product was added to 1 μ l of the TOPO vector, left at room temperature for 5 min, and used to transform *E. coli* Top10 competent cells (Invitrogen). Every positive colony was fully sequenced to discard any mutation due to PCR amplification or TOPO recombination. Then positive colonies were used for plasmid DNA purification.

Destination plasmids pDEST17 (His₆ tag) and pDEST15 (GST tag) were purchased from Invitrogen; plasmid pTH1, for MBP fusions, was a kind gift of Prof. T. Hard (Royal Institute of Technology, Stockholm, Sweden). Transformation of all 29 genes (38 cDNA clones in total, combining partial- and full-length ones) with every tag was attempted. Similar amounts of pENTR/D-TOPO vector and destination vector (300 ng) were mixed with 1 μ l of LR Clonase (Invitrogen) for 1 h at 25 °C. Then 2 μ l of proteinase K were added for 10 min at 37 °C to stop the reaction. In general, we directly transformed BL21(DE3) cells (Edge Biosystems), saving 2 days of work, with acceptable results in terms of efficiency. Only when direct transformation of BL21(DE3) cells failed did we try DH5 α cells (Invitrogen). Positive colonies were confirmed by PCR in all cases. No mutations or deletions were observed in this step.

Protein Expression and Solubility in *E. coli*—Single colonies were grown at 37 °C in LB medium supplemented with 100 μ g/ml ampicillin to midlog phase ($A_{600} = 0.6$ – 0.8) where expression was induced by the addition of 0.4 mM isopropyl 1-thio- β -D-galactopyranoside (Roche Applied Science). After a 3-h induction, cells were harvested by centrifugation and resuspended in 0.05 culture volumes of PBS. Three freeze/thaw cycles followed by sonication on ice were used to disrupt the cells, and the soluble and insoluble fractions were separated by centrifugation. The levels of expression and solubility were analyzed using Coomassie-stained SDS-PAGE and immunoblotting. For SDS-PAGE analysis, 10% SDS-polyacrylamide gels were run, and the resolved proteins were stained with Coomassie Blue G250 (Bio-Rad). For immunoblotting, proteins were transferred onto a Hy-

bond-C nitrocellulose membrane (Amersham Biosciences) and blocked with 3% skimmed milk in PBS containing 0.05% Tween 20 for 1 h at room temperature. Then membranes were incubated with an anti-tag-specific antibody for 2 h at room temperature. Subsequently HRP- or alkaline phosphatase-labeled secondary antibody was added for 1 h at room temperature and visualized using ECL (Amersham Biosciences) or nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate (Bio-Rad) substrate, respectively.

Purification of His₆-, GST-, and MBP-tagged Proteins—Proteins with His₆ fusion were purified by affinity chromatography using a HiTrap™ column (Amersham Biosciences) on an ÄKTA Prime system (Amersham Biosciences) according to standard procedures. When insoluble His₆ proteins were purified, an on-column refolding method was applied: 6 M guanidine hydrochloride, 20 mM Tris-HCl, 0.5 M NaCl, 5 mM imidazole, 1 mM 2-mercaptoethanol, pH 8.0 was used as binding buffer; 6 M urea, 20 mM Tris-HCl, 0.5 M NaCl, 5 mM imidazole, 1 mM 2-mercaptoethanol, pH 8.0 was the solubilization buffer; and 20 mM Tris-HCl, 0.5 M NaCl, 0.5 M imidazole, 1 mM 2-mercaptoethanol, pH 8.0 was the elution buffer.

For GST fusion proteins we used a GSTrap™ column (Amersham Biosciences). The binding buffer was 20 mM sodium phosphate, pH 7.3, with 0.15 M NaCl, and the elution buffer was 50 mM Tris-HCl, pH 8.0, with 10 mM reduced glutathione. In both cases, fusion proteins were pooled and dialyzed against PBS.

Soluble *E. coli*-derived MBP fusion proteins were purified by passing the extracts over an amylose resin column (New England Biolabs) equilibrated in 20 mM Tris HCl, pH 7.4, 200 mM NaCl, 1 mM EDTA and recovered by elution with 10 mM maltose in the same buffer. Fractions were analyzed on 10% SDS-polyacrylamide gels stained with Coomassie Brilliant Blue.

Protein Identification by MALDI-MS—Identity of the recombinant proteins was confirmed by MALDI-TOF mass spectrometry analysis. The protein spots were excised manually and transferred into siliconized 0.5-ml tubes. The gel pieces were washed twice with 50% acetonitrile. Then the gel fragments were placed at 56 °C for 45 min in 10 mM DTT, 55 mM iodoacetamide in 25 mM ammonium bicarbonate in the dark. Approximately 10 μ l of 0.1 μ g/ μ l modified trypsin (Promega) in 25 mM ammonium bicarbonate was added to the gel fragments and incubated overnight at 37 °C. After the supernatant was transferred to an Eppendorf tube, 20 μ l of 50% acetonitrile, 0.1% trifluoroacetic acid was added, and the peptides were further extracted from the gel piece by sonication for 5 min and dried down. Peptides were resuspended in 10 μ l of 33% acetonitrile, 0.1% trifluoroacetic acid.

For MS analysis, a MALDI-TOF mass spectrometer (Autoflex, Bruker Daltonics) was used in positive ion reflector mode. The ion acceleration voltage was 20 kV. Each spectrum was internally calibrated with the masses of two trypsin autolysis products. For peptide mass fingerprint identification, the tryptic peptide mass maps were transferred through the MS BioTools™ program (Bruker Daltonics) as inputs to search Swiss-Prot using Mascot software (Matrix Science). Up to one missed tryptic cleavage was considered, and a mass accuracy of 50 ppm was used for all tryptic mass searches.

Antibody Production—BMP4- (clone ALB190F) and LCN2 (clone HAT265B)-specific monoclonal antibodies were produced as described previously (31) using partial-length MBP-BMP4 and full-length GST-LCN2 antigens, respectively. Hybridoma supernatants were screened initially by ELISA, immunoblotting on transfected cells and cell line and human tissue extracts. Finally they were tested by immunohistochemistry on bone marrow and spleen. Polyclonal antibodies against ANXA3 were produced by immunizing New Zealand White rabbits with the His fusion of the protein.

Immunoblotting Analysis of Tissue Samples—Frozen tissues were washed twice with chilled PBS, and proteins were extracted by son-

Validation of Genomic Targets in Colorectal Cancer

TABLE II

Summary list of the 29 selected target proteins with information regarding the availability of partial- and full-length sequence clones

The three last columns compile the success in the construction of the expression clones. NA, no clone available/found; BC, length of the cDNA sequence below cutoff (30% of the ORF); +, successful expression clone; -, no successful expression clone; F/P, full/partial-length successful construction.

Protein name	Gene symbol	Target clones				
		Molecular mass		MBP fusion proteins	GST fusion proteins	His ₆ fusion proteins
		Full length	Partial length			
Da						
Collagen type V α 2	COL5A2	NA	BC	—	—	—
Integrin α 2	ITGA2	NA	BC	—	—	—
Protein-tyrosine kinase	PTK7	118,391	BC	+	—	+
Ephrin B3	EPHB3	106,520	BC	+	+	+
Sulfatase 1	SULF1	101,026	BC	—	+	+
Solute carrier family 26 member 3	SLC26A3	84,671	BC	+	—	—
Solute carrier family 2 member 1	SLC2A1	NA	BC	—	—	—
Angiotensinogen precursor	AGT	53,154	BC	+	+	+
Carbonic anhydrase IX	CA9	49,698	15,653	—	—	—
Plasminogen activator	PLAU	48,525	BC	+	+	+
Thrombospondin 2	THBS2	NA	47,898	+	+	+
Bone morphogenetic protein 4	BMP4	44,048	23,867	+, + (F, P)	+, +	+, +
Selenoprotein P 1	SEPP1	42,705	BC	+	—	+
Annexin 3	ANXA3	36,375	BC	+	+	+
Osteopontin	SPP1	35,422	21,523	+, +	+, +	+, +
Osteonectin	SPARC	34,632	BC	+	+	+
Insulin-like growth factor-binding protein 3	IGFBP3	31,674	BC	—	—	—
Matrix metalloproteinase 7	MMP7	29,677	19,429	+, +	+, +	+, +
Collagen triple helix repeat-containing 1	CTHRC1	26,224	11,813	+, +	+, +	+, +
Matrix metalloproteinase 11	MMP11	23,628	13,956	—, +	—, +	+, +
Tissue inhibitor of metalloproteinase 1	TIMP1	23,171	BC	+	+	+
Lipocalin 2	LCN2	22,588	15,488	+, +	+, +	+, +
Teratocarcinoma-derived growth factor 1	TDGF1	21,168	BC	+	+	+
Interferon-induced transmembrane protein 3	IFITM3	14,632	14,148	—, +	+, +	+, +
Interferon-induced transmembrane protein 2	IFITM2	14,546	12,452	+, +	+, +	+, +
Interferon-induced transmembrane protein 1	IFITM1	13,964	14,364	—	—	—
GRO3 oncogene	CXCL3	11,342	BC	+	+	+
GRO1 oncogene	CXCL1	7,864	BC	+	+	+
Hypoxia-inducible protein 2	HIG2	6,950	BC	+	+	+
Total	29	25	11	22	20	22
		26 distinct available proteins		19 F	18 F	21 F
				9 P	9 P	9 P

ication in lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 2 mM DTT, 0.1% SDS) three times. Protein concentration was determined using the Protein Assay™ kit (Bio-Rad). Protein extracts from normal and tumoral tissues, from five selected patients from Dukes' stages A-C, were resolved by 12% SDS-PAGE. Proteins were transferred to nitrocellulose membranes (Hybond-C Extra, Amersham Biosciences). After blocking with 5% nonfat milk, membranes were incubated for 90 min with specific antibodies: anti-LCN2 and -BMP4 hybridoma supernatants were used undiluted, -ANXA3 was used at 1:1,000 dilution, -SPARC was used at 1:1,000 dilution, -SPP1 was used at 1:10,000 dilution, -MMP7 was used at 1:2,000 dilution, and -MMP11 was used at 1:200 dilution. Anti-tubulin mouse antibody (Sigma) was used as a loading control at 1:5,000 dilution. Following three washes in PBS containing 0.05% Tween 20, membranes were incubated with a 1:1,000 dilution of either rabbit anti-mouse IgG HRP conjugate (DakoCytomation), or anti-rabbit IgG HRP conjugate (Sigma) for 90 min. Antibody binding was detected using ECL reagent or SuperSignal-Femto (Pierce).

Tissue Microarray Preparation and Analysis—CRC tumoral samples corresponding to the three first stages of Dukes' classification (A, B, and C), their normal mucosa counterparts, and tissue controls such

as tonsil, breast, liver, lung, pancreas, and placenta were used for the preparation of the tissue microarrays. Neoplastic biopsies from two groups of 48 and 49 CRC patients and their non-tumoral mucosal counterparts were collected after tumor resection. Samples were collected and made anonymous by the Tumor Bank Network (CNIO). They were handled according to the ethical and legal standards. Diagnostic paraffin blocks were selected on the basis of the availability of suitable formalin-fixed paraffin-embedded tissue, containing enough remaining tissue for a minimum of 60 sections. Histological confirmation of CRC was achieved in all cases by central review using standard tissue sections, and the most tumor-rich areas were marked in the paraffin blocks. We used a tissue arrayer device (Beecher Instruments) to construct two microarray blocks containing a total of 97 samples, including replicates for some of the tumors and representing different locations of the neoplasia along the colonic region. Protein expression was graded 1–3 in the microarray (1, <10%; 2, 10–50%; 3, >50% of stained tumoral specific location). Two selected 1-mm-diameter cylinders from two different areas were included in each case along with several controls to ensure the quality, reproducibility, and homogenous staining of the slides. The arrays were incubated with antibodies against ANXA3, BMP4, and LCN2 at dilu-

Validation of Genomic Targets in Colorectal Cancer

tions of 1:500, 1:25, and 1:2, respectively. Anti-SPARC (1:25 dilution) was from Novocastra, anti-SPP1 (1:750) was from Abcam, and anti-MMP7 (1:200) and anti-MMP11 (1:50 dilution) were from Chemicon. Ki67 (1:50 dilution, DakoCytomation) was included as a control for proliferation. Specific binding was followed by incubation with prediluted commercial anti-mouse/rabbit IgG conjugated with biotin. Visualization of specific interactions was monitored by using the EnVision HRP system (Dako) following the manufacturer's instructions.

RESULTS

Identification of Up-regulated Genes in CRC by DNA Microarray Analysis—Twenty-two paired samples of CRC patients were analyzed for differential gene expression by using a cDNA microarray (CNIO Oncochip). After one class SAM analysis, a total of 1,182 probes were found to be up-regulated at a delta value threshold corresponding to a q-value <1% (data not shown). At this lower limit, because the q-value established the lowest false discovery rate, the number of false positives in this set of selected genes was lower than 12. In addition, we established an upper limit of 0.6 in the 22-patient averaged log₂Ratio to filter out genes showing low global transcription increases in the tumors. In this way, 371 probes showed more than 1.5-fold change in CRC tumoral samples. The 371 probes matched 337 known genes and 13 hypothetical proteins (Supplemental Table S1).

Prediction of Surface and Extracellular Proteins as Putative Targets—Further selection criteria were based on subcellular location and gene ontology characteristics of the gene products. We preferentially selected those genes between the 100 most up-regulated genes among the 337 found genes with only a few exceptions (*SLC2A1*, *HIG2*, *CA9*, *EPHB3*, and *ITGA2*) that were placed between positions 100 and 200 (Supplemental Table S1). After bioinformatic analysis, a total of 29 proteins were selected for expression (Table II). Only proteins with cell surface accessibility or extracellular location were selected because these proteins would be accessible for antibody-mediated diagnosis or therapy. The range of molecular masses varied between 129 kDa for THBS2 down to 6.9 kDa for hypoxia-inducible gene 2 (*HIG2*). Many of those proteins do not have commercially available antibodies against them.

Expression of Partial-length cDNA Clones with Three Different Tags—Collecting and validating the cDNA clones for every gene was a time-consuming and lengthy process. The initial idea was to use the same cDNA clones from the IMAGE collection used for printing the Oncochip microarray as the source of ORFs for protein expression. Unfortunately only 12 of 29 clones covered at least 30% of the ORF sequence length. Only nine of 12 could be used for gene expression; the other three (*ANX3*, *IFITM1*, and *TIMP1*) consistently showed deletions after PCR amplification and cloning into pENTR/D-TOPO vector. A total of 27 destination clones were prepared after recombination of those nine entry clones with the three destination vectors. The efficiency of recombination was lower and more variable when *E. coli* BL21(DE3) cells were

used for direct transformation as compared with DH5 α cells: 70% versus >95% at the first attempt. However, for high throughput purposes, the overall time savings justifies direct transformation of BL21(DE3). The results for protein expression and solubility in *E. coli* are summarized in Table II and shown as Coomassie-stained gels (Fig. 1A) and immunoblotting analysis (Fig. 1B).

Soluble products were obtained for seven of nine MBP fusion proteins (78%), six GST fusion proteins (67%), and two His-tagged proteins (22%). These results confirm previous findings that MBP is a more efficient solubilizing agent for expression and recovery of soluble proteins in *E. coli* than GST or His₆ tags. The low efficiency of His₆ was surprising. No clone was soluble with the three tags, but all of them were correctly expressed and solubilized with at least one tag. In addition, BMP4, IFITM2, IFITM3, MMP11, and SPP1 were soluble with two tags. In summary, by combining the three tags all these protein fragments could be recovered in a soluble form.

Expression of Full-length cDNA Clones with the Three Tags—To get full-length ORFs, we used either the MGC (24 of 29 clones (82%)) or donations from other laboratories. Three cDNAs (*COL5A2*, *ITGA2*, and *SLC2A1*) were not found. A total of 26 clones containing full-length cDNAs were used for expression. After removal of the leader sequences, coding regions were PCR-amplified and cloned into the pENTR/D-TOPO vector with a variable efficiency between 15 and 100%. No entry clones were obtained for three genes (*CA9*, *IGFBP3*, and *IFITM1*). The entry clones were subsequently recombined with the same three destination vectors as above. Data for expression of the proteins in *E. coli* BL21(DE3) and solubility properties are summarized in Table III. Coomassie Blue and immunoblotting analyses are shown in Fig. 2, A and B, respectively. No destination clones were obtained for many of the large proteins (>50 kDa), suggesting an increasing difficulty for recombination. The efficiency of cloning ORF PCR products up to 2 kb is about 90% for the Gateway system but decreases considerably for larger clones (32).

In our hands, none of the proteins above 50 kDa, except SULF1, were soluble with any of the different tags when expressed in *E. coli*. MBP fusion proteins were soluble more often than fusions with GST or His₆. Interestingly all the proteins below 42.7 kDa (corresponding to SEPP1) were soluble when fused to MBP (13 of 19 (68%)). AGT and PLAU expression seems to correspond to the MBP tag alone. For GST, 5 of 18 (28%) were soluble. AGT and TIMP1 products seem to be mainly GST. In the case of His₆, solubilization was observed in 7 of 22 cases (32%), mostly for proteins with molecular masses below 40 kDa.

Purification of Fusion Proteins and MALDI-TOF Analysis—Fusion proteins with the three tags were purified for immunization purposes. Results are shown in Fig. 3. Soluble MBP fusion proteins were purified by using amylose resins in a batch procedure. A characteristic of those elutions was the

Validation of Genomic Targets in Colorectal Cancer

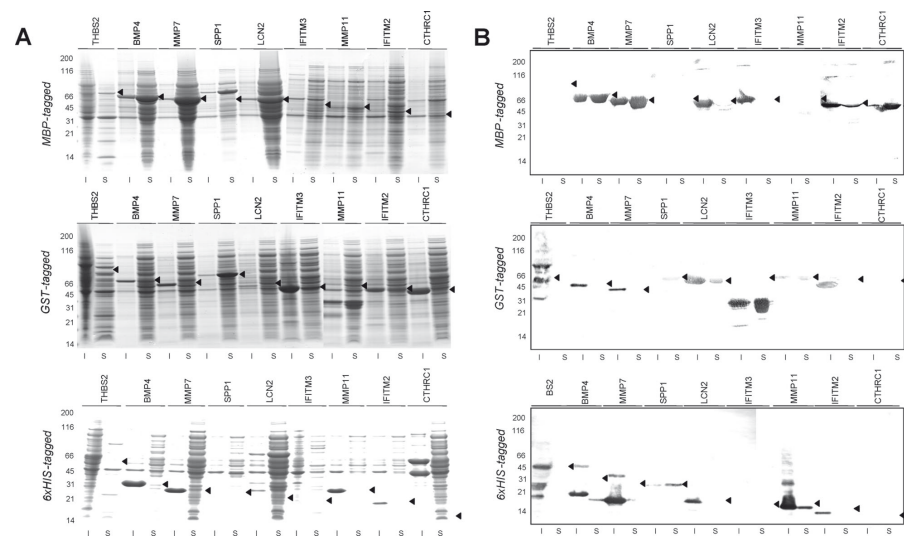


FIG. 1. Expression-solubility profile of the selected partial-length cDNAs fused to three different affinity tags: MBP, GST, and His₆. These cDNAs were transferred into each of the three expression vectors by recombinational cloning and transformed into BL21(DE3) cells. Bacteria were grown, induced, and lysed as described under "Experimental Procedures." The insoluble fraction for each expression was dissolved in a volume equal to that of the soluble fraction and loaded into 10% SDS-PAGE gels. **A**, Coomassie-stained gels showing soluble (S) and insoluble (I) fractions of the corresponding recombinant protein identified at the top of every two lanes. The upper panel corresponds to MBP-tagged proteins, the middle panel corresponds to GST-tagged proteins, and the lower panel corresponds to the His₆ fusions. **B**, immunoblotting analyses using antibodies against each tag as indicated at the left of each panel. Marker sizes are given in the left side of each panel. The expected size of each protein is indicated by an arrow on the right side of the S lanes.

recurrent presence of minor bands underneath the correct one that probably arise from degradation fragments of the MBP. Although this fact might be disturbing for other uses of the protein, it is not the case for the preparation of antibodies as all of them are derived from the fusion proteins. Similar findings were obtained for purified GST-tagged proteins, which also showed the presence of several lower molecular weight bands. With respect to the His₆ fusion proteins, five His₆-tagged proteins were purified under denaturing conditions followed by a refolding on-column. In general, although His fusions were recovered as soluble proteins in only a few cases, they usually yielded purer proteins in a single step.

Every fusion protein (partial- and full-length constructions) prepared with any of the three tags was subsequently analyzed by mass spectrometry to confirm its identity. Nineteen of 20 analyzed full-length expressed proteins were correctly identified as expected except for one case (SLC26A3) that corresponded to a non-visible band product. The whole set of partial-length products (nine of nine) were also positively validated.

Validation Studies by Immunoblotting—Seven proteins (ANXA3, BMP4, LCN2, SPARC, SPP1, MMP7, and MMP11)

were initially selected as potential CRC markers for their novelty and relationship to some interesting pathways in tumor development and progression. To check whether those proteins were truly overexpressed in the samples, protein extracts from normal and tumoral tissues from six patients representing stages A–C of CRC were resolved by SDS-PAGE, blotted onto nitrocellulose membranes, and incubated with these antibodies. Fig. 4 shows the results obtained. ANXA3, MMP7, and SPARC antibodies were particularly efficient in tumoral versus normal differentiation with no apparent preference for either early or advanced stages. BMP4 recognized preferentially A and C stages, although normal samples also showed significant expression. A trend was observed for LCN2 with respect to the progression stage showing an increase in the expression for the more advanced stages corresponding to Dukes' B and C series. MMP11 allowed for a significant discrimination between normal and tumoral samples, especially in stage C samples. The only molecule that could not be validated by immunoblotting as being differentially expressed in carcinoma samples was SPP1, which showed an erratic pattern along the whole series of samples. Tubulin was used as a positive control for normalization.

Validation of Genomic Targets in Colorectal Cancer

TABLE III
Expression and solubility levels of 26 target proteins when cloned as three different gene fusions

No entry clones were obtained for three genes (CA9, IFITM1, and IGFBP3). From the remaining 23 genes with verified expression clones, tagged products matching 21 different proteins were expressed (full- and/or partial-length sequence). NA, no clone available; F, full-length cDNA sequence; P, partial-length cDNA sequence; E, expression; S, solubility; U, unsuccessful cloning attempt.

Protein name	Gene symbol	Ontology	Full or partial length	Molecular mass	MBP fusion proteins		GST fusion proteins		His ₆ fusion proteins		Mass spectrometry analysis
					E ^a	S ^b	E ^a	S ^b	E ^a	S ^b	
Da											
Protein-tyrosine kinase	PTK7	Signal transduction	F	118,391	—	—	U	U	—	—	
			P	NA							
Ephrin B3	EPHB3	Angiogenesis	F	106,520	++	0	+	0	+	0	+
			P	NA							
Sulfatase 1	SULF1	Apoptosis	F	101,026	U	U	+	0	+++	+	+
			P	NA							
Solute carrier family 26 member 3	SLC26A3	Sulfate transporter	F	84,671	—	—	U	U	U	U	—
			P	NA							
Angiotensinogen precursor	AGT	Blood pressure	F	53,154	*	*	*	*	+	+	+
			P	NA							
Carbonic anhydrase IX	CA9	Hypoxia marker	F	49,698	U	U	U	U	U	U	
			P	15,653	U	U	U	U	U	U	
Plasminogen activator	PLAU	Growth factor	F	48,525	+	+	++	+	+++	0	+
			P	NA							
Thrombospondin 2	THBS2	Angiogenesis	F	47,898	—	—	++	++	+	—	+
			P	NA							
Bone morphogenetic protein 4	BMP4	Growth factor, cytokine	F	44,048	+++	0	++	+	+	—	+
			P	23,867	+++	+++	+++	++	+++	0	+
Selenoprotein P 1	SEPP1	Extracellular matrix	F	42,705	—	—	U	U	+++	++	+
			P	NA							
Annexin 3	ANXA3	Calcium ion binding	F	36,375	++	+++	+++	++	+++	+	+
			P	NA							
Osteopontin	SPP1	Growth factor	F	35,422	—	—	+	—	—	—	+
			P	21,523	+++	+++	+++	+++	—	—	+
Osteonectin	SPARC	Growth factor	F	34,632	++	++	—	—	++	+	+
			P	NA							
Insulin-like growth factor-binding protein 3	IGFBP3	Growth factor	F	31,674	U	U	U	U	U	U	
			P	NA							
Matrix metalloproteinase 7	MMP7	Extracellular matrix	F	29,677	+++	+++	+++	0	+++	0	+
			P	19,429	+++	+++	+++	0	*	*	+
Collagen triple helix repeat-containing 1	CTHRC1	Extracellular matrix	F	26,224	+++	+++	+++	+	+++	0	+
			P	11,813	++	+++	+++	0	+	0	+
Matrix metalloproteinase 11	MMP11	Extracellular matrix	F	23,628	U	U	U	U	++	+++	
			P	13,956	++	+++	++	++	*	*	+
Tissue inhibitor of metalloproteinase 1	TIMP1	Growth factor	F	23,171	++	++	—	—	—	—	+
			P	NA							
Lipocalin 2	LCN2	Oncogene	F	22,588	+++	+++	+++	++	+++	++	+
			P	15,488	++	+++	++	+	+	+++	+
Teratocarcinoma-derived growth factor 1	TDGF1	Growth factor	F	21,168	++	+++	+++	+	++	—	+
			P	NA							
Interferon-induced transmembrane 3	IFITM3	Cell cycle	F	14,632	U	U	+++	+	+++	+	+
			P	14,148	++	++	+++	+	—	—	
Interferon-induced transmembrane 2	IFITM2	Cell cycle	F	14,546	+	+	+++	0	—	—	+
			P	12,452	++	+	+++	++	+++	++	
Interferon-induced transmembrane 1	IFITM1	Cell cycle	F	13,964	U	U	U	U	U	U	
			P	14,364	U	U	U	U	U	U	
GRO3 oncogene	CXCL3	Growth factor, cytokine	F	11,342	++	++	+++	+	—	—	+
			P	NA							
GRO1 oncogene	CXCL1	Growth factor, cytokine	F	7,864	+++	+++	+++	++	—	—	+
			P	NA							
Hypoxia-inducible protein 2	HIG2	Hypoxia marker	F	6,950	—	—	+++	0	—	—	+
			P	NA							

^a Expression levels given as: +++ = strong band on SDS-PAGE gel; ++ = medium; + = visible band; — = no visible band; * = incorrect size.

^b Solubility given as: +++ = most of the protein in soluble fraction; ++ = roughly 50% in soluble fraction; + = minority in soluble fraction; 0 = nothing in soluble fraction; — = no expression.

Downloaded from www.mcponline.org at CNIC on January 17, 2008

Validation of Genomic Targets in Colorectal Cancer

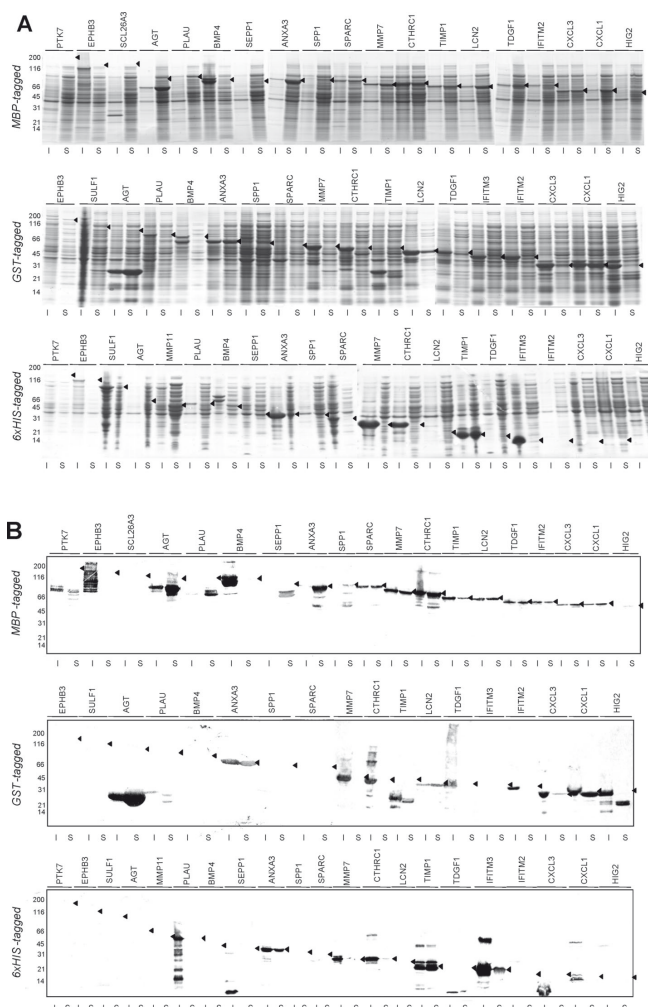


FIG. 2. Expression-solubility profile of the selected full-length cDNAs fused to three different affinity tags: MBP, GST, and His₆. Growing and expression conditions were the same as those described for partial-length proteins. **A**, Coomassie Blue-stained gels of soluble (S) and insoluble (I) fractions of the recombinant full-length protein named at the top of each two lanes. **B**, detection of fusion proteins by immunoblotting with antibodies against the corresponding tags: MBP, upper panel; GST, middle panel; and His₆, lower panel. Marker sizes are given in the left side of each panel. The expected size of each protein is indicated by an arrow on the right side of the S lanes.

Applications of the Selected Proteins in CRC Diagnosis—Protein expression in tumoral tissues was confirmed by immunohistochemistry on two distinct CRC-specific tissue microarrays containing 97 different paired human tissues (Fig. 5). Annexin A3 staining showed a clear preference for the tumoral tissues; an increase in the intensity of the signal was noticed for two-thirds of the cases (63%). In most cases there was a specific reactivity with epithelial cells surrounding the

crypts and villi (51%), whereas the remaining tissues exhibited staining of stromal cells. Although specific staining for ANXA3 was detected in the cytoplasm, the clearest reaction was observed in the cellular membrane. Normal samples barely showed annexin A3 expression in epithelial cells.

BMP4 was detected throughout the cytoplasm of tumoral cells as well as in the extracellular matrix surrounding the mucosal cells, displaying a scattered, sometimes granular

Validation of Genomic Targets in Colorectal Cancer

FIG. 3. Purification of fusion proteins by affinity chromatography. Coomassie Blue-stained SDS-PAGE showing the purification of the selected proteins accordingly with their tags. MBP fusions were purified with amylose resins, GST fusions were purified with glutathione columns, and His₆ (6xHis) fusions were purified with Ni²⁺ columns. Marker sizes are given in the left side of the figure. The expected size of each purified protein is indicated by an arrow on the right side of the lanes. Partial-length proteins are labeled with (p) before the protein name.

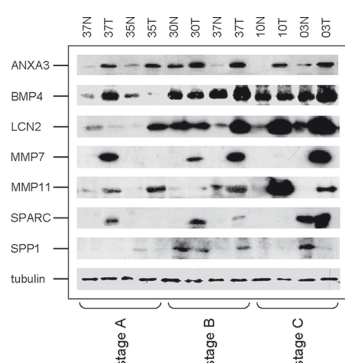
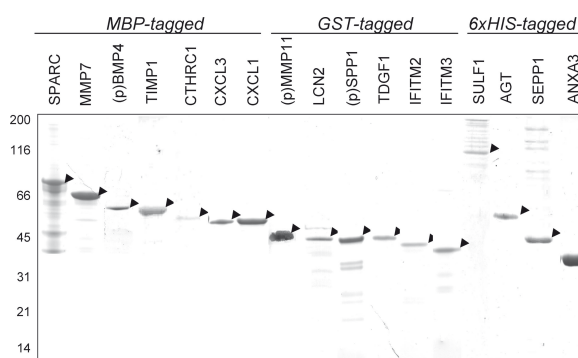


FIG. 4. Immunoblotting analysis of selected colorectal cancer targets in tissue samples. The study was performed for those proteins for which we developed antibodies (ANXA3, LCN2, and BMP4) as well as commercial antibodies for SPARC, SPP1, and MMP11. Anti-tubulin antibody was used as a control. Conventional one-dimensional 10% SDS-PAGE gels were run with protein extracts from paired normal (N) and tumoral (T) tissues from six CRC patients (two from each of the Dukes stages A–C). Proteins were transferred onto nitrocellulose membranes and incubated with specific antibodies raised against the target proteins synthesized in our approach. Reactivity was revealed by chemiluminescence (ECL) or SuperSignal Femto (Pierce).

staining in agreement with the condition of BMP4 as a secreted protein. In a minor percentage of the cells, nuclei were also marked; this is in accordance with the reported translocation of BMP4 for the activation of the Smad pathway. Normal tissues gave a weak staining confined to some areas in the stroma and not in the epithelial cells.

Within CRC samples, there were a high number of epithelial cells expressing LCN2. The staining was cytoplasmic with a clear reinforcement of the membrane signal. The LCN2-pos-

itive cells were homogeneously distributed in the luminal area of the microvilli with a distinctive staining of their apical epithelium. Expression of LCN2 was markedly detected in the sigmoid colon (78% of the cases) and sigmoid-rectum (83%) where 57% of the positive cases were intermediately or strongly stained. With respect to other colon locations (such as cecum; ascending, transverse, and descending colon; and rectum), the reactivity and intensity varied among the cases, ~50% of which stained positively. No LCN2 protein expression was observed in the normal mucosa tissues tested.

In the case of MMP7, the staining of the tumoral cells was cytoplasmic, with a strong signaling marking the Golgi apparatus, and preferentially located in the apical regions of the neoplasia, next to the lumen, in focal locations at the end of the crypts or in areas of incipient metastasis. Apart from these focal nodes, MMP7 antibody stained the stroma surrounding the epithelial cells to some extent even in normal mucosa, giving the average tissue slice a weak background. Associated to MMP7 and sharing a similar pattern, SPP1 staining appeared in only a small percentage of the cases (13%) and usually was associated with focal expression near the cavities of the glands or blood capillaries. The signal in positive tumoral cells is visible throughout the cytoplasm, and because SPP1 is a secreted protein, a faint expression is also found in the extracellular matrix even in normal tissues in some cases.

MMP11 showed a juxtatumoral staining of stromal cells in CRC tissues that is probably due to the role of MMP11 in invasiveness. No staining was observed in normal tissues. A similar observation was made for SPARC for which the stroma subjacent to the cells was clearly stained in neoplastic samples, was substantially reduced in dysplasias, but was not at all present in hyperplasias and normal tissues. The SPARC labeling in juxtatumoral stromal cells developed as a strong and homogeneous cytoplasmic staining.

As a marker for proliferation, we used Ki67. A strong staining of the carcinoma cells was observed in the tumoral tissues

Downloaded from www.mcponline.org at CNIC on January 17, 2008

Validation of Genomic Targets in Colorectal Cancer

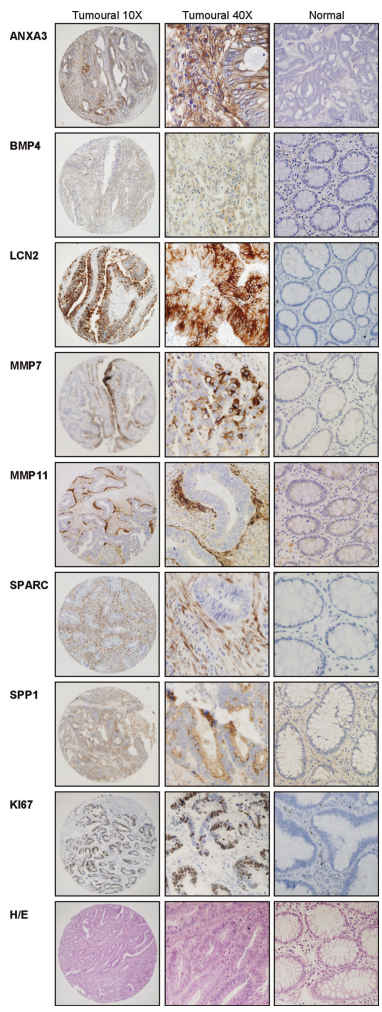


FIG. 5. Tissue microarray analysis of the selected targets in colorectal cancer. Selected images of colorectal cancer tissue microarray elements representing immunohistochemical staining of the selected seven proteins (ANXA3, BMP4, LCN2, MMP7, MMP11, SPARC, and SPP1) at different magnifications (10× and 40×) in tumoral tissues and normal tissues are shown. Control Ki67 preferentially marked those cells undergoing proliferation, a distinctive label for neoplastic tissues. A hematoxylin/eosin (H/E) staining of section controls is also included.

reflecting their proliferative activity.

DISCUSSION

After gene expression analysis of CRC samples using DNA microarrays, a collection of up-regulated genes was investigated for their ability as biomarkers for CRC. There were several objectives in this study: first, to investigate the possibilities to translate as fast as possible these genomic targets into soluble proteins; second, to test the recombinant proteins for the preparation of monoclonal and/or polyclonal antibodies; third, to check these antibodies for the validation of the genomic results; and fourth, to investigate and characterize these potential new markers for CRC. The selection of up-regulated CRC genes was non-biased, that is, independent of gene size, location, and biochemical function. We did not exclude large molecular weight proteins, integral membrane proteins, or secreted proteins. On the contrary, we specifically focused on these (membrane and secreted) proteins as more likely to be biomarkers of interest and targets for antibody diagnostics and therapy.

The initial idea was to use for protein expression the cDNA clones used regularly for the printing of DNA chips. However, these clones consisted of long untranslated regions, and none of these “chip” clones contained a full-length cDNA of the coding region. In many cases the ORF did not even reach the 30% that we had set up as our lower limit for expression experiments. This process of sequence verification for each clone already consumed a significant amount of time. As an alternative, full-length clones were obtained from the MGC repository or from colleagues after intensive search. For 30 genes the whole collection process was lengthy and time-consuming, making the retrieval of cDNA clones one of the bottlenecks of the whole process. Availability of public repositories containing full-length and sequence-verified ORF clones in recombinational cloning systems such as Gateway or Creator would be extremely beneficial for a high throughput expression project (33).

To speed up the cloning process, we selected TOPO-Gateway vectors in which the topoisomerase enzyme has been attached to the free ends of the vector. TOPO vectors present several advantages over pDONR vectors. They need shorter oligonucleotides for recombination; therefore, fewer foreign residues are introduced, and the cloning efficiency is higher. About 90% of pENTR constructions were generated at the first attempt. In the remaining 10% of the clones, small deletions were observed that obliged us to sequence every clone, slowing down the process. For instance, the entry clone for IFITM1 showed an internal 5-nucleotide deletion, which aborted the cloning process at an early stage. Another limitation of the Gateway system is the need to perform up to three transformation steps in *E. coli*, one for the entry clone and two more for the destination vectors, first in DH5 and then in BL21 strains. With our procedure, we skipped one step by transforming *E. coli* BL21 directly with the LR Clonase prod-

Validation of Genomic Targets in Colorectal Cancer

uct. Ninety-five percent of the attempted transformations succeeded at the first round, and the rest were obtained in a second attempt.

At the present time, it is impossible to predict the results of protein expression. There were some previous attempts to predict expression probabilities based on the characteristics of the protein family and the Pfam domains with little success (34). In the present study, our observation is that, by using this high throughput approach, only proteins below 50 kDa were expressed as soluble proteins with at least one of the tags; this constitutes a severe limitation for these genomic projects. MBP tag was the best choice for an increased solubility of the fusion proteins. The solubility decreased significantly in proportion to the size of the protein. The fact that MBP fusion proteins are soluble does not mean that they are correctly folded or functionally active (35, 36). However, it is evident that for many applications such as antibody production these soluble proteins represent a great advantage due to easier handling and a lower toxicity for the host animals. Moreover the possibility of developing antibodies against natural conformational epitopes is much higher using soluble proteins as shown by the reactivity obtained by the monoclonal antibodies developed against LCN2 and BMP4.

When proteins were fused to GST or His₆ tag only 27.7 and 13.6% were recovered in a soluble form, respectively. In a previous study (20), a similar approach was followed to express 27 small human protein (below 20 kDa), obtaining similar solubility results with MBP and thioredoxin tags. However, thioredoxin is not appropriate for affinity purification and would require a second affinity tag for purification. Braun *et al.* (19) expressed and purified 32 human proteins using the same three tags plus calmodulin-binding protein. They described similar results of protein recovery for MBP and GST and did not provide details about solubility of the proteins. We believe that our procedure still needs some improvements in automation and especially in the purification steps, making it more amenable to high throughput, by using chemical lysis and microplate batch purification procedures as reported previously (19).

In summary, from the initially selected 29 genes, only 21 distinct proteins were finally expressed with, at least, one different fusion protein. With respect to the full- or partial-length sequence option, seven proteins were achieved in both formats, and 12 of them were obtained as complete sequence products, whereas the remaining two could only be obtained as protein fragments. According to these considerations and limitations, either new tags should be developed for a more efficient recovery of soluble proteins, or alternative eukaryotic systems should be optimized for high throughput approaches (37).

Recent advances in proteomic research underscore the increasing need for high affinity antibodies, which are still generated with lengthy, low throughput antibody production techniques. Other alternative routes such as phage display are currently being tested. Regarding antibody production, in

our hands, no significant differences in time were observed between preparing monoclonal or polyclonal antibodies. Moreover the performance of the polyclonal antibody against ANXA3 equals or outperforms the use of monoclonal antibodies for BMP4 or LCN2. Therefore, the final method of choice should be based in terms of availability or accessibility to antibody production facilities if pure antigens are on hand. Probably for a first screening, polyclonal antibodies may represent a cheaper alternative.

To test the usefulness of this approach for marker validation, we chose seven proteins for antibody production and/or validation. These seven potential markers were selected for their novelty and/or relationship to some interesting pathways in CRC tumor development, invasiveness, and progression. ANXA3 encodes a member of the annexin family. Members of this calcium-dependent phospholipid-binding protein family play a role in the regulation of cellular growth and in signal transduction pathways. There are no previous reports of association of this protein to tumoral processes. Lipocalin 2, also known as neutrophil gelatinase-associated lipocalin, is a 25-kDa protein stored in specific granules of the human neutrophil (38). It might function as a modulator of inflammation. The LCN2 gene is highly homologous to the mouse oncogene 24p3. Potential cis-acting elements include binding sites for transcription factors GATA-1 and PU.1 and NFκB. LCN2 was found to be highly expressed after malignant transformation of the breast, lung, colon, and pancreatic epithelia by immunohistochemistry (39). More recently, it has been described that LCN2 decreases the invasiveness and metastasis of Ras-transformed cells (40). BMP4 is a member of the bone morphogenetic protein family, which is part of the transforming growth factor-β superfamily. This superfamily includes large families of growth and differentiation factors. Recent experiments revealed that the oncogenic allele of β-catenin is required for BMP4 expression and secretion by human cancer cells and that BMP4 is overexpressed and secreted by human colon cancer cells with mutant adenomatous polyposis coli (APC) genes. These data identify the presence of regulatory interactions between the Wnt and BMP signaling pathways in cancer pathogenesis, providing an intriguing connection between the sporadic and inherited forms of a common human malignancy (41). BMP4 also promotes melanoma cell invasion and migration (42). SPARC, also called osteonectin, is a multifunctional matricellular glycoprotein that has been associated with impaired tumor growth, antiangiogenic properties, apoptosis induction, and changes in the extracellular matrix (43). SPP1, also called osteopontin, is overexpressed in a wide number of tumors, inducing various cellular signaling events leading to the activation of various kinases, urokinase plasminogen activator, and matrix metalloproteases (44). It plays a crucial role in tumor progression. MMP7 and MMP11 are members of the MMP family, which is involved in the breakdown of extracellular matrix in normal physiological processes, such as em-

Validation of Genomic Targets in Colorectal Cancer

bryonic development, reproduction, and tissue remodeling, as well as in disease processes, such as arthritis and metastasis. MMP7 is elevated in several human cancers and seems to be related to CRC liver metastases (45). MMP11 protein is involved in the pathway of colorectal cancer development in females, distal locations, infiltrative growth patterns, and microsatellite stability (46).

Our results have proven that there is a good correlation between the transcriptomic analysis and the protein expression data for the four genes tested. There was a clear overexpression in most cases; overexpression was particularly significant in the cases of ANXA3, SPARC, MMP7, and LCN2 for making a distinction between tumoral and normal tissues. Remarkably about 73% of all the cases showed LCN2 reactivity (they were more strongly staining Dukes' B stages). MMP11 also showed a tendency to be found in late stages. The staining pattern of some of these partners, mainly in the stromal tissue, stresses the importance of considering the whole tissue for evaluations because some proteins are differentially expressed in stromal cells surrounding and adjacent to regions of diseased epithelium that correlate with tumor progression (47).

In summary, we have demonstrated the feasibility of this approach for a relatively quick expression and validation of a high number of genomic targets. We have identified some major bottlenecks and suggested ways to overcome them. Although this preliminary characterization was only restricted to seven targets, it is clear that approaches like this may yield relevant biological information about the neoplastic processes and lead to the characterization of potentially interesting markers in a quite straightforward manner for early diagnosis or individual prognosis.

Acknowledgments—We thank Manuel Morente from the CNIO Tissue Bank Network and Jose Solano from the Pathology Department of the Hospital of Cartagena for cooperation in the collection of CRC samples. We thank A. Nuñez, S. Romero, F. Fernandez, and P. Alfonso for technical assistance and the staff of the monoclonal antibody unit of the CNIO for assistance in the preparation of the different antibodies. We also thank the technical staff of the histology and immunohistochemistry unit of the CNIO for assistance in the preparation of the tissue microarrays and immunohistochemical staining.

* This work was supported in part by Spanish Ministry of Science Grant BIO2003-01481. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

¶ Both authors contributed equally to this work.

¶ Supported by a contract grant from the Comunidad Autónoma de Madrid.

|| A Becas de formación en investigación (BEFI) fellow of the Instituto de Salud Carlos III.

§§ To whom correspondence should be addressed: Protein Technology Unit, Biotechnology Programme, Centro Nacional de Investigaciones Oncológicas, Melchor Fernandez Almagro 3, 28029 Madrid,

Spain. Tel.: 34-91-2246920; Fax: 34-91-2246972; E-mail: icasal@cnio.es.

REFERENCES

1. Fearon, E. R. (1996) Colorectal cancer: molecular genetic studies and their future clinical applications. *Med. Pediatr. Oncol. Suppl.* **1**, 35–40
2. Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470
3. Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675–1680
4. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511
5. DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14**, 457–460
6. Williams, N. S., Gaynor, R. B., Scoggin, S., Verma, U., Gokaslan, T., Simmang, C., Fleming, J., Tavara, D., Frenkel, E., and Becerra, C. (2003) Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference. *Clin. Cancer Res.* **9**, 931–946
7. Zou, T. T., Selaru, F. M., Xu, Y., Shustova, V., Yin, J., Mori, Y., Shibata, D., Sato, F., Wang, S., Orlau, A., Deacu, E., Liu, T. C., Abraham, J. M., and Meltzer, S. J. (2002) Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. *Oncogene* **21**, 4855–4862
8. Birkenkamp-Demtroder, K., Christensen, L. L., Olesen, S. H., Frederiksen, C. M., Laiho, P., Aaltonen, L. A., Laurberg, S., Sorensen, F. B., Hagemann, R., and Orntoft, T. F. (2002) Gene expression in colorectal cancer. *Cancer Res.* **62**, 4352–4363
9. Notterman, D. A., Alon, U., Sierk, A. J., and Levine, A. J. (2001) Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res.* **61**, 3124–3130
10. Welsh, J. B., Zarrinkar, P. P., Sapinoso, L. M., Kern, S. G., Behling, C. A., Monk, B. J., Lockhart, D. J., Burger, R. A., and Hampton, G. M. (2001) Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1176–1181
11. Welsh, J. B., Sapinoso, L. M., Su, A. I., Kern, S. G., Wang-Rodriguez, J., Moskaluk, C. A., Frierson, H. F., Jr., and Hampton, G. M. (2001) Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res.* **61**, 5974–5978
12. Hellstrom, I., Raycraft, J., Hayden-Ledbetter, M., Ledbetter, J. A., Schummer, M., McIntosh, M., Drescher, C., Urban, N., and Hellstrom, K. E. (2003) The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. *Cancer Res.* **63**, 3695–3700
13. Alfonso, P., Nuñez, A., Madoz-Gurpide, J., Lombardia, L., Sanchez, L., and Casal, J. I. (2005) Proteomic expression analysis of colorectal cancer by two-dimensional differential gel electrophoresis. *Proteomics* **5**, 2602–2611
14. Friedman, D. B., Hill, S., Keller, J. W., Merchant, N. B., Levy, S. E., Coffey, R. J., and Caprioli, R. M. (2004) Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics* **4**, 793–811
15. Olesen, S. H., Christensen, L. L., Sorensen, F. B., Cabezon, T., Laurberg, S., Orntoft, T. F., and Birkenkamp-Demtroder, K. (2005) Human FK506 binding protein 65 is associated with colorectal cancer. *Mol. Cell. Proteomics* **4**, 534–544
16. Tracey, L., Villuendas, R., Ortiz, P., Dopazo, A., Spiteri, I., Lombardia, L., Rodriguez-Peralto, J. L., Fernandez-Herrera, J., Hernandez, A., Fraga, J., Dominguez, O., Herrero, J., Alonso, M. A., Dopazo, J., and Piris, M. A.

Validation of Genomic Targets in Colorectal Cancer

- (2002) Identification of genes involved in resistance to interferon- α in cutaneous T-cell lymphoma. *Am. J. Pathol.* **161**, 1825–1837
17. Strausberg, R. L., Feingold, E. A., Klausner, R. D., and Collins, F. S. (1999) The mammalian gene collection. *Science* **286**, 455–457
 18. Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Altschul, S. F., Zeeberg, B., Buetow, K. H., Schaefer, C. F., Bhat, N. K., Hopkins, R. F., Jordan, H., Moore, T., Max, S. I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A. A., Rubin, G. M., Hong, L., Stapleton, M., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Scheetz, T. E., Brownstein, M. J., Ustin, T. B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S. S., Loquellano, N. A., Peters, G. J., Abramson, R. D., Mullahy, S. J., Bosak, S. A., McEwan, P. J., McKernan, K. J., Malek, J. A., Gunaratne, P. H., Richards, S., Worley, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hulyk, S. W., Villalón, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Young, A. C., Shevchenko, Y., Bouffard, G. G., Blakesley, R. W., Touchman, J. W., Green, E. D., Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M., Butterfield, Y. S., Krzywinski, M. I., Skalska, U., Smallus, D. E., Schnerch, A., Schein, J. E., Jones, S. J., and Marra, M. A. (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16899–16903
 19. Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E., and LaBaer, J. (2002) Proteome-scale purification of human proteins from bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 2654–2659
 20. Hammarstrom, M., Helligren, N., van Den Berg, S., Berglund, H., and Hard, T. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Sci.* **11**, 313–321
 21. Walhout, A. J., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., van den Heuvel, S., and Vidal, M. (2000) GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* **328**, 575–592
 22. Hartley, J. L., Temple, G. F., and Brasch, M. A. (2000) DNA cloning using in vitro site-specific recombination. *Genome Res.* **10**, 1788–1795
 23. Uhlen, M., Bjorling, E., Agaton, C., Szgyarto, C. A., Amini, B., Andersen, E., Andersson, A. C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergstrom, K., Brumer, H., Cerjan, D., Ekstrom, M., Elbeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Bjorklund, M. G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kamp, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundberg, J., Magnusson, K., Malm, E., Nilsson, P., Odling, J., Oksvold, P., Olsson, I., Oster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Skolleremo, A., Steen, J., Stenvall, M., Sterky, F., Stromberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Walden, A., Wan, J., Wernerus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L. L., Hober, S., and Ponten, F. (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* **4**, 1920–1932
 24. Tyers, M., and Mann, M. (2003) From genomics to proteomics. *Nature* **422**, 193–197
 25. Tracey, L., Spiteri, I., Ortiz, P., Lawler, M., Piris, M. A., and Villuendas, R. (2004) Transcriptional response of T cells to IFN- α : changes induced in IFN- α -sensitive and resistant cutaneous T cell lymphoma. *J. Interferon Cytokine Res.* **24**, 185–195
 26. Martinez, N., Camacho, F. I., Algara, P., Rodriguez, A., Dopazo, A., Ruiz-Ballesteros, E., Martin, P., Martinez-Climent, J. A., Garcia-Conde, J., Menarguez, J., Solano, F., Mollejo, M., and Piris, M. A. (2003) The molecular signature of mantle cell lymphoma reveals multiple signals favoring cell survival. *Cancer Res.* **63**, 8226–8232
 27. Tracey, L., Villuendas, R., Dotor, A. M., Spiteri, I., Ortiz, P., Garcia, J. F., Peralto, J. L., Lawler, M., and Piris, M. A. (2003) Mycosis fungoides shows concurrent deregulation of multiple genes involved in the TNF signaling pathway: an expression profile study. *Blood* **102**, 1042–1050
 28. Lin, Y. M., Furukawa, Y., Tsunoda, T., Yue, C. T., Yang, K. C., and Nakamura, Y. (2002) Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. *Oncogene* **21**, 4120–4128
 29. Smyth, G. K., and Speed, T. (2003) Normalization of cDNA microarray data. *Methods* **31**, 265–273
 30. Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 5116–5121
 31. Roncador, G., Garcia, J. F., Maestre, L., Lucas, E., Menarguez, J., Ohshima, K., Nakamura, S., Banham, A. H., and Piris, M. A. (2005) FOXP3, a selective marker for a subset of adult T-cell leukaemia/lymphoma. *Leukemia* **19**, 2247–2253
 32. Marsischky, G., and LaBaer, J. (2004) Many paths to many clones: a comparative look at high-throughput cloning methods. *Genome Res.* **14**, 2020–2028
 33. Brizuela, L., Richardson, A., Marsischky, G., and Labaer, J. (2002) The FLEXGene repository: exploiting the fruits of the genome projects by creating a needed resource to face the challenges of the post-genomic era. *Arch. Med. Res.* **33**, 318–324
 34. Klein, J., and Dhurjati, P. (1995) Protein aggregation kinetics in an *Escherichia coli* strain overexpressing a *Salmonella typhimurium* CheY mutant gene. *Appl. Environ. Microbiol.* **61**, 1220–1225
 35. Sachdev, D., and Chirgwin, J. M. (1998) Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin. *Protein Expr. Purif.* **12**, 122–132
 36. Martinez-Torrecuadrada, J. L., Romero, S., Nunez, A., Alfonso, P., Sanchez-Céspedes, M., and Casal, J. I. (2005) An efficient expression system for the production of functionally active human LKB1. *J. Biotechnol.* **115**, 23–34
 37. Albala, J. S., Franke, K., McConnell, I. R., Pak, K. L., Folta, P. A., Rubinfeld, B., Davies, A. H., Lennon, G. G., and Clark, R. (2000) From genes to proteins: high-throughput expression and purification of the human proteome. *J. Cell. Biochem.* **80**, 187–191
 38. Cowland, J. B., and Borregaard, N. (1997) Molecular characterization and pattern of tissue expression of the gene for neutrophil gelatinase-associated lipocalin from humans. *Genomics* **45**, 17–23
 39. Friedl, A., Stoesz, S. P., Buckley, P., and Gould, M. N. (1999) Neutrophil gelatinase-associated lipocalin in normal and neoplastic human tissues. Cell type-specific pattern of expression. *Histochem. J.* **31**, 433–441
 40. Hanai, J., Mammoto, T., Seth, P., Mori, K., Karumanchi, S. A., Barasch, J., and Sukhatme, V. P. (2005) Lipocalin 2 diminishes invasiveness and metastasis of Ras-transformed cells. *J. Biol. Chem.* **280**, 13641–13647
 41. Kim, J. S., Crooks, H., Dracheva, T., Nishanian, T. G., Singh, B., Jen, J., and Waldman, T. (2002) Oncogenic β -catenin is required for bone morphogenetic protein 4 expression in human cancer cells. *Cancer Res.* **62**, 2744–2748
 42. Rothhammer, T., Poser, I., Soncin, F., Bataille, F., Moser, M., and Bosserhoff, A. K. (2005) Bone morphogenic proteins are overexpressed in malignant melanoma and promote cell invasion and migration. *Cancer Res.* **65**, 448–456
 43. Chlenski, A., Liu, S., Guerrero, I. J., Yang, Q., Tian, Y., Salwen, H. R., Zage, P., and Cohn, S. L. (2006) SPARC expression is associated with impaired tumor growth, inhibited angiogenesis and changes in the extracellular matrix. *Int. J. Cancer* **118**, 310–316
 44. Rangaswami, H., Bulbule, A., and Kundu, G. C. (2006) Osteopontin: role in cell signaling and cancer progression. *Trends Cell Biol.* **16**, 79–87
 45. Zeng, Z. S., Shu, W. P., Cohen, A. M., and Guillem, J. G. (2002) Matrix metalloproteinase-7 expression in colorectal cancer liver metastases: evidence for involvement of MMP-7 activation in human cancer metastases. *Clin. Cancer Res.* **8**, 144–148
 46. Skoglund, J., Emterling, A., Arbm, G., Anglard, P., and Sun, X. F. (2004) Clinicopathological significance of stromelysin-3 expression in colorectal cancer. *Oncology* **67**, 67–72
 47. Mueller, M. M., and Fusenig, N. E. (2004) Friends or foes—bipolar effects of the tumour stroma in cancer. *Nat. Rev. Cancer* **4**, 839–849

Downloaded from www.mcponline.org at CNIO on January 17, 2008



A continuación se incluyen los resúmenes de otras publicaciones científicas en las que ha participado el doctorando durante el desarrollo de la tesis doctoral:

Clin. Cancer Res. 2005 Sep 1; 11 (17): 6280-6290.

Targeting the extracellular domain of fibroblast growth factor receptor 3 with human single-chain Fv antibodies inhibits bladder carcinoma cell line proliferation.

Martínez-Torreuadrada, J., Cifuentes, G., **López-Serra, P.**, Saenz, P., Martínez, A. and Casal, J. I.

PURPOSE: Previous gene expression studies have shown that fibroblast growth factor receptor 3 (FGFR3) is overexpressed in early stages of bladder cancer. To study the potential use of therapeutic antibodies against FGFR3, we have produced a collection of human single-chain Fv (scFv) antibody fragments by using phage display libraries. **EXPERIMENTAL DESIGN:** Two "naïve" semi-synthetic human scFv libraries were used to select antibodies against the extracellular domain of FGFR3alpha(IIIc). The reactivity of the selected scFvs with a recombinant FGFR3 was characterized by an enzyme immunoassay and surface plasmon resonance analysis and with RT112 bladder carcinoma cells by a fluorescence-activated cell sorter. The capacity of the selected scFvs to block RT112 cell proliferation was determined. **RESULTS:** We have isolated six human scFv antibody fragments directed against FGFR3. These human scFvs specifically bound FGFR3, but not the homologous molecule FGFR1. Biacore analysis was used to determine the affinity constants, which ranged from 12 to 40 nmol/L. Competition analysis showed that the FGF9 ligand was able to block the binding of two scFvs, 3C and 7D, to FGFR3, whereas FGF1 only blocked 7D. Immunoprecipitation and flow cytometric analysis confirmed the specificity of the antibodies to native membrane FGFR3. Two scFvs, 3C and 7D, gave an strong immunofluorescence staining of RT112 cells. Moreover, they recognized equally well wild-type and mutant FGFR3 containing the activating mutation S249C. Furthermore, they blocked proliferation of RT112 cells in a dose- and FGF-dependent manner. **CONCLUSION:** Our results suggest that these human anti-FGFR3 scFv antibodies may have potential applications as antitumoral agents in bladder cancer.

Mol. Cancer Ther. 2008 April; 7 (4): 862-873.

Antitumor activity of fibroblast growth factor receptor 3-specific immunotoxins in a xenograft mouse model of bladder carcinoma is mediated by apoptosis

Jorge L. Martínez-Torrecuadrada, Lawrence H. Cheung, **Paula López-Serra**, Rodrigo Barderas, Marta Cañamero, Sergio Ferreiro, Michael G. Rosenblum and J. Ignacio Casal.

ABSTRACT: Human single-chain Fv directed against fibroblast growth factor receptor 3 (FGFR3) have been shown to block proliferation of RT112 bladder carcinoma cells in vitro. Here, we examined the ability of the recombinant gelonin toxin (rGel) to enhance this inhibitory effect in vitro and in vivo on the bladder cancer cell line RT112 and the

corresponding xenografts. Immunotoxins were genetically engineered by fusing FGFR3-specific Fv fragments (3C) to the NH2 terminus of rGel and expressed as a soluble

protein in *Escherichia coli*. The 3C/rGel fusion construct showed an IC₅₀ of 200 nmol/L against log-phase RT112 cells compared with 1,500 nmol/L for free rGel. Immunofluorescence studies showed that the 3C/rGel construct internalized rapidly into the cytoplasm of RT112 cells within 1 h of exposure. The mechanism of immunotoxin-induced cell death was found to be mediated by apoptosis. RT112 tumor xenografts in severe combined immunodeficient mice treated with 50 mg/kg 3C/rGel exhibited considerable growth delay relative to control tumors and a significant reduction of 55% to 70% in mean tumor size. Immunohistochemical analysis showed that tumors from mice treated with 3C/rGel displayed considerable apoptotic damage compared with control groups. Subcellular location of FGFR3 in immunotoxin-treated tumors indicated

a translocation of FGFR3 to the nuclear membrane in contrast to tumors from saline-treated controls. These results show that FGFR3-driven immunotoxins may be an effective therapeutic agent against human bladder and other tumor types overexpressing FGFR3.